

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:18:51 ; Search time 8.0038 Seconds  
(without alignments)  
2502.854 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1881  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgacgtag 1041

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US09751299/runat\_26042006\_090036\_18677/app\_query.fasta\_1  
-DB=PIR -QFMT=faстан -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US09751299 @CGN 1\_1\_77 @runat\_26042006\_090036\_18677 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	869	46.2	356	2	A47181	nitrilase (EC 3.5.
2	832.5	44.3	366	2	A45070	nitrilase - Rhodoc
3	825.5	43.9	383	2	A43470	aliphatic nitrilas
4	616	32.7	349	1	A28658	nitrilase (EC 3.5.
5	592.5	31.5	354	2	JC4212	nitrilase (EC 3.5.
6	584.5	31.1	346	2	S77025	nitrilase (EC 3.5.
7	562	29.9	368	2	JQ1613	cyanide hydratase
8	468.5	24.9	346	2	T49148	nitrilase (EC 3.5.
9	467	24.8	346	2	T49147	nitrilase (EC 3.5.
10	464	24.7	339	2	S31969	nitrilase (EC 3.5.
11	463	24.6	339	2	T52259	nitrilase (EC 3.5.
12	461	24.5	346	2	S22398	nitrilase (EC 3.5.
13	456	24.2	339	2	T52262	nitrilase (EC 3.5.
14	420	22.3	362	2	T52266	nitrilase-like pro

15	411	21.9	348	2	T03739	nitrilase (EC 3.5.
16	404.5	21.5	349	2	T03736	nitrilase (EC 3.5.
17	403.5	21.5	305	2	T27679	probable nitrilase
18	399.5	21.2	355	2	T52265	nitrilase (EC 3.5.
19	226	12.0	529	2	T45134	hypothetical prote
20	224.5	11.9	199	2	S50363	nitrilase homolog
21	223	11.9	1106	2	J00405	hypothetical 119.5
22	201.5	10.7	1460	1	EDBEIF	immediate-early pr
23	199.5	10.6	579	2	AH3225	amidohydrolase (lm
24	198	10.5	262	2	C75051	hydrolase related
25	196.5	10.4	385	2	G87519	hypothetical prote
26	196.5	10.4	1791	2	T02345	hypothetical prote
27	196	10.4	262	2	C71109	hypothetical prote
28	195	10.4	494	2	G84348	hypothetical prote
29	192	10.2	1106	2	J00405	hypothetical 119.5
30	191.5	10.2	1791	2	T02345	hypothetical prote
31	188.5	10.0	437	2	C39135	hypothetical prote
32	183.5	9.8	414	2	G86301	arginine/serine-ri
33	182	9.7	318	2	T48563	hypothetical prote
34	178.5	9.5	891	2	G84693	probable prolina-r
35	178.5	9.5	1122	2	T14180	exit protein - Myc
36	176.5	9.4	857	2	T05352	hypothetical prote
37	175.5	9.3	494	2	G84348	hypothetical prote
38	173	9.2	376	2	C75580	adenine deaminase-
39	173	9.2	376	2	C75580	adenine deaminase-
40	170.5	9.1	529	2	T45134	hypothetical prote
41	170	9.0	1037	2	B87254	translational initia
42	169	9.0	580	2	T43481	probable mucin DKF
43	168	8.9	358	1	WMBE38	infected cell prot
44	166.5	8.9	660	1	QOBE3	BHLFI protein - hu
45	165.5	8.8	682	2	T28899	hypothetical prote

ALIGNMENTS

RESULT 1

A47181  
nitrilase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis

C:Species: Alcaligenes faecalis

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #ext\_change 31-Dec-2004

C:Accession: A47181; S13860

R:Kobayashi, M.; Izui, H.; Nagasawa, T.; Yamada, H.

Proc. Natl. Acad. Sci. U.S.A. 90, 247-251, 1993

A:Title: Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole

A:Reference number: A47181; MUID:93126352; PMID:8419330

A:Contents: JMJ3

A:Accession: A47181

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-356 <KOB>

A:Cross-references: UNIPROT:P20960; UNIPARC:UPI0000033514; GB:D13419; NID:9216202; PIDN:K

A:Note: sequence extracted from NCBI backbone (NCBIN:122081, NCBIPI:122082)

R:Nagasawa, T.; Manger, J.; Yamada, H.

Eur. J. Biochem. 194, 765-772, 1990

A:Title: A novel nitrilase, arylacetoneitrilase, of Alcaligenes faecalis JMJ3. Purification

A:Reference number: S13860; MUID:91099356; PMID:2269298

A:Accession: S13860

A:Status: preliminary

A:Molecule type: protein

A:Residues: 1-33 <NAG>

A:Cross-references: UNIPARC:UPI0000175E31

C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

C:Keywords: hydrolase

Alignment Scores:  
Pred. No.: 6.26e-49  
Score: 869.00  
Percent Similarity: 65.3%  
Best Local Similarity: 50.6%  
Query Match: 46.2%  
DB: 2  
US-09-751-299-1 (1-1041) x A47181 (1-356)

QY 25 CGCGCGCGCGCGATGCAGCGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAGAA 84  
Db 8 ArgAlaAlaAlaValGlnAlaAlaSerProAsnTyrAspLeuAlaThrGlyValAspLys 27  
QY 85 GCGATCGGCTGATCGAGCAGCGCGCGCAAGCAGCAGTCGCGCTGATCGCATCCAGAG 144  
Db 28 ThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCysAspLeuIleValPheGlyGlu 47  
QY 145 ACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGCGCGCTTGCGCATGCGC 204  
Db 48 ThrTyrLeuProGlyTyrProPheHisValIlePheGlyAlaProAlaIleTrpSerLeuLys 67  
QY 205 TTCGTCCAGCGCTATTTCGAGAAATTCGCTCGTCGCGCGCAGCAAGCAGTCGCGCGCTG 264  
Db 68 TyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGluPheGlnArgIle 87  
QY 265 GCGGATGCGCGCGCGCGCGCGCATGTCATGCTGGCGCGGCTATAGCGAGCGCGCGGC 324  
Db 88 AlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSerGluArgSerGly 107  
QY 325 GCGAGCTCTATATGGCGCAGCGCATCTTCGCGCGCGCGCATGCGCATGCGCGCGC 384  
Db 108 GlySerLeuTyrLeuGlyGlnCysLeuIleAspAspLysGlyGlnMetLeuIleTrpSerArg 127  
QY 385 CGCAAGCTCAAGCCTACCCATGCGCGGAGCGCAACCGGTTCGGCGAGGAGAGCGGACCAT 444  
Db 128 ArgLysLeuLysProThrHisValaGluArgThrValPheGlyGluGlyTyrAlaArgAsp 147  
QY 445 CTCGCGGTGCAGATACCGGCATCGCGCGCGCTCGCGCGCGCTGTGTTGCTGGGAGACATC 504  
Db 148 LeuIleValSerAspThrGluLeuGlyArgValaGlyAlaLeuCysTyrTrpGluHisLeu 167  
QY 505 CAGCCATTGTGAAATAAGCCATGTACGCGCGCGCGCAAGCAAGCTCCAGTCCGCTCGTG 564  
Db 168 SerProLeuSerLysTyrAlaLeuTyrSerGlnHisGluAlaIleHisIleAlaIleAlaTrp 187  
QY 565 CCGAGCTTCAGCCTCTATCGCGCATGCGCTATGCGCTCGGACCGCGAGTCAATACCGCC 624  
Db 188 ProSerPheSerLeuTyrSerGluGlnAlaHisAlaLeuSerAlaLysValAsnMetAla 207  
QY 625 GCAAGCCAGATGTAAGCGGTGAGCGCGCGCTCTACGTGCTGCGCTGCGCGACCGCTT 684  
Db 208 AlaSerGlnIleTyrSerValaGluGlyGlnCysPheThrIleAlaAlaSerSerValVal 227  
QY 685 TCGCGGAGATGATCAAGTATGTTGGATACGCCGCAAGAGATGTTCTCAAGGCC 744  
Db 228 ThrGlnGluThrLeuAspMetLeuGluValaGlyGluHisAsnAlaSerLeuLeuLysVal 247  
QY 745 GCGGCGGTTTGGCATGATTTTCGCGCGCGCGCGCGCGCTGCGCGCGCGCTGCGCGCG 804  
Db 248 GlyGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAlaProTyrLeuPro 267  
QY 805 GAGACCGAAGAGGAGTGTGTCGCGCGCATATGACCTCGGCATGATGCGCTGGCGCAAG 864  
Db 268 HisAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluGluIleAlaPheAlaLys 287  
QY 865 GCGGCGCGCATCGCGCGCGCAATTCACGCGCGCGCGCAAGTCAAGCGCGCTGCTGGAT 924  
Db 288 AlaIleAsnAspProValaGlyHisTyrSerLysProGluAlaThrArgLeuValLeuAsp 307  
QY 925 -----CGACGTCCGCGCGCAAGCGTGTGTCAGCGCTTGATCGCGCATTCGAACCGCAA 975  
Db 308 LeuGlyHisArgGluProMetThrArgValHisSerLysSerValIleGlnGlu----- 325  
QY 976 AACGAGGACAAGGCGCGCGCGCGCTGCGCGGTGCGCGGAAAGCGCGCGCGCGCG 1035  
Db 326 -----GluAlaProGluProHisValaGlnSerThrAlaAlaProValAla 340

RESULT 2  
A45070  
nitrilase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous

C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A45070  
R/Kobayashi, M.; Kameda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
J. Biol. Chem. 267, 20746-20751, 1992  
A/Title: Nitrlase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the  
A/Reference number: A45070; MUID:93015976; PMID:1400390  
A/Contents: J1  
A/Accession: A45070  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-366 <KOB>  
A/Cross-references: UNIPROT:Q03217, UNIPARC:UPI000016FF36; GB:D11425; NID:g216933; PIDN:1  
A/Note: sequence extracted from NCBI backbone (NCBIN:116039, NCBIPI:116040)  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)

Alignment Scores:  
Pred. No.: 1.48e-46 Length: 366  
Score: 832.50 Matches: 166  
Percent Similarity: 65.7% Conservative: 43  
Best Local Similarity: 52.2% Mismatches: 108  
Query Match: 44.3% Indels: 1  
DB: 2 Gaps: 1

US-09-751-299-1 (1-1041) x A45070 (1-366)

QY 1 ATGTCGAGACCCCATGACCAAGTATCGCGCGCGGTGCAGCGCGCGCGGTGTTCTCCTC 60  
Db 1 MetValGluTyrThrAsnThrPheLysValaAlaValaGlnAlaGlnProValTrpPhe 20  
QY 61 GATTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCGCAAGCAGAC 120  
Db 21 AspAlaAlaLysThrValaAspLysThrValSerIleIleAlaGluAlaAlaArgAsnGly 40  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATCCCGCTATCCCTTTGGATATGCTG 180  
Db 41 CysGluLeuValaAlaPheProGluValPheIleProGlyTyrProTyrHisIleTrpVal 60  
QY 181 GCGCGCGCGGCTGGCGCATG--CGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTG 237  
Db 61 AspSerProLeuAlaGlyMetAlaLysPheAlaValaArgTyrHisGluAsnSerLeuThr 80  
QY 238 CGCGCAGCAAGCAGTGGCAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCATGCTC 297  
Db 81 MetAspSerProHisValaGlnArgLeuLeuAspAlaAlaArgAspHisAsnIleAlaVal 100  
QY 298 GTGCGCGCTATACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTATATGAGCGCATCTTCGGC 357  
Db 101 ValValaGlyIleSerGluArgAspGlyGlySerLeuTyrMetThrGlnLeuValIleAsp 120  
QY 358 CCCGATGGCATCTGATCGCGCGCGCGCGCGCGCGCGCGCAAGCTCAAGCTTACCATCGGAGCGCAC 417  
Db 121 AlaAspGlyGlnLeuValaAlaArgArgArgLysLeuLysProThrHisValaGluArgSer 140  
QY 418 GTGTCGCGGAGGAGACGCGCAGCATCTCGCGGTGCAGATACCGGCATCGCGCGCGCTC 477  
Db 141 ValTyrGlyGluGlyAsnGlySerAspIleSerValTyrAspMetProPheAlaArgLeu 160  
QY 478 GCGCGCTCTGTTGCTGGAGCAGCATCCAGCATGTCGAAATAAGCCATGATGAGCGCGCC 537  
Db 161 GlyAlaLeuAsnCysTrpGluHisPheGlnThrLeuThrLysTyrAlaMetTyrSerMet 180  
QY 538 GACGAAGAGTCCACGTGCGGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTAT 597  
Db 181 HisGluGlnValHisValaIleSerTrpProGlyMetSerLeuTyrGlnProGluValPro 200  
QY 598 GCGCTCGGACCGGAGGTGAATAAGCGCGCGCAAGCAGATCAAGCGGTGAGGCGCGCTGC 657  
Db 201 AlaPheGlyValaAspAlaGlnLeuThrAlaThrArgMetTyrAlaLeuGluGlnThr 220  
QY 658 TACGTCTGCGCGTGTGCG 717  
Db 221 PheValValaCysThrThrGlnValValThrProGluAlaHisGluPhePheCysAspAsn 240



QY 718 CCCGACAGAGATGTTCTCAAGCGCGCGGCTTTTGGCATGATTTTGGCCCGAC 777  
Db 241 AspGlnArgLysLeuIleGlyArgGlyGlyPheAlaArgIleIleGlyProAsp 260  
QY 778 GGGCGCGCCCTGGCCGAGCGCTCCCGAGACCGAAGAGGACTGCTGTCGCCGATATC 837  
Db 261 GlyArgAspLeuAlaThrProLeuAlaGluAspGluGlyIleLeuTyrAlaAspIle 280  
QY 838 GACCTCGGCATGATCGCGTTGGCCCAAGCGCGCGCGATCCGGCGGCCACTATTACGG 897  
Db 281 AspLeuSerAlaIleThrLeuAlaLysGlnAlaAlaAspProValGlyHisTyrSerArg 300  
QY 898 CCCGACGTAACGGCGCTGCTGTCGATCGACGTCGCCGCCAACGCGCTGTCACG 951  
Db 301 ProAspValLeuSerLeuAsnPheAsnGlnArgHisThrThrProValAsnThr 318

RESULT 3

A43470  
aliphatic nitrilase - Rhodococcus rhodochrous  
C:Species: Rhodococcus rhodochrous  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C:Accession: A43470  
R:Kobayashi, M.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
Biochemistry 31, 9000-9007, 1992  
A:Title: Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase  
A:Reference number: A43470, MUID:93003039; PMID:1390687  
A:Contents: K22  
A:Accession: A43470  
A:Status: preliminary  
A:Molecule type: DNA; protein  
A:Residues: 1-383 <KOB>  
A:Cross-references: UNIPROT:Q02068; UNIPARC:UPI0000130516; GB:D12583; NID:g216931; PIDN:  
A:Note: sequence extracted from NCBI backbone (NCBIN:114184, NCBI P:114185)  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

Alignment Scores:  
Pred. No.: 4.23e-46 Length: 383  
Score: 825.50 Matches: 169  
Percent Similarity: 60.9% Conservative: 49  
Best Local Similarity: 47.2% Mismatches: 125  
Query Match: 43.9% Indels: 15  
DB: 2 Gaps: 6

US-09-751-299-1 (1-1041) x A43470 (1-383)

QY 4 TCGAGAGCCCATGACGAAGTATCGCGC-----GCCGCGTGCAGCGCGCGCG 51  
Db 3 SerAsnProGluLeuLysTyrThrGlyLysValLysValAlaThrValGlnAlaGluPro 22  
QY 52 GTGTTCTCGATCTGCACCGCACAGTCAGAAAGCGATCGCGCTGATCGACGAGCGCGC 111  
Db 23 ValIleLeuAspAlaAspAlaThrIleAspLysAlaIleGlyPheIleGluGluAlaAla 42  
QY 112 AAGCAGGAGTGCCTGATCGCATTCCTCCAGAGACTTGATTCCTGCTATCCCTTTGG 171  
Db 43 LysAsnGlyAlaGluPheLeuAlaPheProGluValTyrIleProGlyTyrProTyrTrp 62  
QY 172 ATATGGCTGGCGCGCGCGCTTGGGCAATCGC--TTCGTCCAGCGCTATTTCGAGAA 228  
Db 63 AlaTyrIleGlyAspValLysTyrPheAlaValSerAspPheIleProLysTyrHisGluAsn 82  
QY 229 TCGCTCGTGCAGCGCGCAGCAGCAGTGCAGCGCCCTGGCGGATGCGCGCGCGCGCAGCGC 288  
Db 83 SerLeuThrLeuGlyAspAspArgMetArgArgLeuGlnLeuAlaAlaArgGlnAsnAsn 102  
QY 289 ATGCATGTCGTGCGCGCTATAGCGAGCGCGCGCGCGCAGCCTTATATGGCCAGCGC 348  
Db 103 IleAlaLeuValMetGlyTyrSerGlyLysAspGlyAlaSerArgTyrLeuSerGlnVal 122  
QY 349 ATCTTGGCGCGCGATGCGATCTGATCGCGCGCGCGCGCAGCTCAAGCTTACCATGCG 408  
Db 123 PheIleAspGlnAsnGlyAspIleValAlaAsnArgArgLysLeuLysProThrHisVal 142

QY 409 GAGCGACCGTGTTCCGGCAGGAGACGGACCCATCTCGCGGTGCACGATACGCCATC 468  
Db 143 GluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeuThrHisAspPheGlyPhe 162  
QY 469 GGGCGCTCGCGCGCTGTGCTGGGAGACATCCAGCCTTGTGAAATACGCCATG 528  
Db 163 GlyArgValGlyGlyLeuAsnGlyTyrGlnHisPheGlnProLeuSerLysTyrMetMet 182  
QY 529 TACGCCCGCAGCAGACAGGTCCAGTTCGCGCTGAGCGCGAGCTTCAACCTTATCGCGC 588  
Db 183 TyrSerLeuAsnGluGlnIleHisValAlaSerTyrProAlaMetPheAlaLeuThrPro 202  
QY 589 ATGGCTATGCGCTCGGACCGGAGCTCAATACCGCGCAGCAGATCTACCGGTCGAG 648  
Db 203 AspValHisGlnLeuSerValGluAlaAsnAspThrValThrArgSerTyrAlaIleGlu 222  
QY 649 GCGCGCTCTACGCTGCGCTGCGCTGCGCGACCGGTTTCCGCGAGATGATCAAGGTATTG 708  
Db 223 GlyGlnThrPheValLeuAlaSerThrHisValIleGlyLysAlaThrGlnAspLeuPhe 242  
QY 709 GTGGATACGCGCGAC--AAGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATT 765  
Db 243 AlaGlyAspAspAlaLysArgAlaLeuLeuProLeuGlyGlnGlyTyrAlaArgIle 262  
QY 766 TTCGCGCGCGACGCGCGCGCGCTGCGCGCGCGCTCCCGGAGACCGAAGAGGACTGCTG 825  
Db 263 TyrGlyProAspGlyLysSerLeuAlaGluProLeuProGluAspAlaGluGlyLeuLeu 282  
QY 826 GTCGCGCATATCGACCTCGGCATGATCGCGTGGCCAAAGCGCGCGCGCGCGC 885  
Db 283 TyrAlaGluLeuAspLeuGluGlnIleIleLeuAlaLysAlaAlaAspProAlaGly 302  
QY 886 CACTATTCACGCGCGCGACGTAAACGCGCTGCTGTCGATCGACGT-----CCGCGC 936  
Db 303 HisTyrSerArgProAspValLeuSerLeuLysIleAspThrArgAsnHisThrProVal 322  
QY 937 CAACGCGTGTACAGCTTGATGCGGCATTCGAACCGCAAAAGAGACAAGCGGACGCG 996  
Db 323 Gln---TyrIleThrAlaAspGlyArgThrSerLeuAsnSerAsnSerArgValGluAsn 341  
QY 997 CCGCGCTGCGCGCTGCTGCGC-----GAAAGCGCGCGCGCGCGC 1035  
Db 342 TyrArgLeuHisGlnLeuAlaAspIleGluLysTyrGluAsnAlaGluAlaAla 359

RESULT 4

A28658  
nitrilase (EC 3.5.5.1) - Klebsiella ozaenae  
C:Species: Klebsiella ozaenae  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C:Accession: A28658  
R:Stalker, D.M.; Malyj, L.D.; McBride, K.E.  
J. Biol. Chem. 263, 6310-6314, 1988  
A:Title: Purification and properties of a nitrilase specific for the herbicide bromoxynil  
A:Reference number: A28658; MUID:88198177; PMID:2834373  
A:Accession: A28658  
A:Molecule type: DNA  
A:Residues: 1-349 <STA>  
A:Cross-references: UNIPROT:P10045; UNIPARC:UPI000013051A; GB:J03196; NID:g149174; PIDN:  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Alignment Scores:  
Pred. No.: 1.74e-32 Length: 349  
Score: 616.00 Matches: 139  
Percent Similarity: 60.1% Conservative: 64  
Best Local Similarity: 41.1% Mismatches: 113  
Query Match: 32.7% Indels: 22  
DB: 1 Gaps: 9

US-09-751-299-1 (1-1041) x A28658 (1-349)

QY 16 ACGAATATCGCGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTGACCGCACA 75  
Db 16 ACGAATATCGCGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTGACCGCACA 75

[illegible]

Gene 161, 15-20, 1995  
A/Title: Aliphatic nitrilase from a soil-isolated *Comamonas testosteroni* sp.: Gene cloning  
A/Reference number: JC4212; MUID:95369726; PMID:7642130  
A/Accession: JC4212  
A/Molecule type: DNA  
A/Residues: 1-354 <LEV>  
A/Cross-references: UNIPROT:Q59329; UNIPARC:UPI00000B98BB; GB:L32589; NID:g1082008; PIDN  
A/Accession: PC4056  
A/Molecule type: protein  
A/Residues: 154-162;295-302;323-340 <LE2>  
A/Cross-references: UNIPARC:UPI0000175E32; UNIPARC:UPI0000175E33; UNIPARC:UPI0000175E34  
C/Comment: This enzyme is active on adiponitrile and cyanovaleric acid.  
C/Genetics:  
A/Gene: nitrA  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
F;163/Active site: Cys #status predicted

**Alignment Scores:**

Pred. No.:	5.88e-31	Length:	354
Score:	592.50	Matches:	137
Percent Similarity:	54.0%	Conservative:	51
Best Local Similarity:	39.4%	Mismatches:	133
Query Match:	31.5%	Indels:	27
DB:	2	Gaps:	7

US-09-751-299-1 (1-1041) x JC4212 (1-354)

QY 31 GCGCGGTGCAGGCGCGCGCGGTGTCTTCGATCTCGACCGCACAGTCGAGAAAGCATC 90  
 |||||  
 10 AlaAlaValGlnAlaAlaProValPheMetAsnLeuGlnAlaThrValAspLysThrCys 29  
 QY 91 GGCCTGATCGAGCAGCGCGCGCCCAAGCAGCAGCAGCGCGCTGATCGCATTTCCAGAGCTTGG 150  
 |||||  
 30 LysLeuIleAlaGlnAlaAlaSerMetGlyAlaLysValIleGlyPheProGlnAlaPhe 49  
 QY 151 ATTCGCGGCTATCCCTTTGGATATGGCTGGC-----GCGCGGCTTGG 195  
 |||||  
 50 IleProGlyTyrProTyrTrpIleTrpThrSerAsnMetAspPheThrGlyMetMetTrp 69  
 QY 196 GGCATGCGGCTTCGTCACGCGCTATTTCGAGAAATTCGCTCGTGGCGGCGGACAGCAGTGG 255  
 :::  
 70 AlaVal-----LeuPheLysAsnAlaIleGlnIleProSerLysGlnVal 84  
 QY 256 CAGGCGCTGCGGATGCGGCGCGCGCGCGCATGTCATGTCGTGGCGGCTATAGCAG 315  
 ||| ::::: |||||  
 85 GlnGlnIleSerAspAlaAlaLysLysAsnGlyValTyrValCysValSerValSerGln 104  
 QY 316 CGCGCGGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGCGCCCGAGTGCATCTGATC 375  
 :::  
 105 LysAspAsnAlaSerLeuTyrLeuThrGlnLeuTrpPheAspProAsnGlyAsnLeuIle 124  
 QY 376 GCCGCGCGCGCAAGCTCAAGCCTACCCATCGCGAGCGCACCGTGTTCGCGCAGAGAGAC 435  
 |||||  
 125 GlyLysHisArgLysPheLysProThrSerSerGluArgAlaValTrpGlyAspGlyAsp 144  
 QY 436 GGCAGCCCATCTCGCGGTGCACGATACCGGCATCGGGCGCTCGCGCGCTCTTGTCTGG 495  
 |||||  
 145 GlySerMetAlaProValPheLysThrGlnTyrGlyAsnLeuGlyGlyLeuGlnCysTrp 164  
 QY 496 GAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCAGCAGAACAGTCCACGTC 555  
 |||||  
 165 GluHisAlaLeuProLeuAsnIleAlaIleMetGlySerLeuAsnGlnGlnValHisVal 184  
 QY 556 GCGTCGTGGCCGAGCTTC-----AGCCTTATCGCGGCGCATGGCC 594  
 |||||  
 185 AlaSerTrpProAlaPheValProLysGlyAlaValSerSerArgValSerSerVal 204  
 QY 595 TATGCGCTCGGACCGGAGGTCAATACCGCGCGCAGCCAGATCTTACCGCGTGCAGGGCGGC 654  
 |||  
 205 CysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyrAlaIleSerAsnGln 224  
 QY 655 TGCTACGTGCTGCGGTGTCGCGCGCAGCGGTTTCGCGGAGATGATCAAGTATTGGTGGAT 714

Db 225 ValTyrValIleMetSerThrAsnLeuValGlyGlnAspMetIleAspMetIleGlyLys 244  
Qy 715 ACGCCCGACAAGAGATGTTCTCAAGCGCGCGGCTTTTGCATGATTTTCGGGCC 774  
Db 245 AspGluPheSerLysAsnPheLeuProLeuGlySerGlyAsnThrAlaIleIleSerAsn 264  
Qy 775 GACGCGCGCGCTGCGAGCCGCTCCGAGACCGAAGAGGAGTCTGTCGCCGAT 834  
Db 265 ThrGlyGluIleLeuAlaSer---IleProGlnAspAlaGluGlyIleAlaValAlaGlu 283  
Qy 835 ATCGACCTCGGCATGATCGCGTTGGCCAGCGCGCGCGATCCGCGCGGCCACTATTCA 894  
Db 284 IleAspLeuAsnGlnIleIleTyrGlyLysTrpLeuLeuAspProAlaGlyHisTyrSer 303  
Qy 895 CGCGCCGACGTAAACGCGGCTGCTGCTGCATCGACGT-----CCGAGCCCAACGCGTC 945  
Db 304 ThrProGlyPheLeuSerLeuThrPheAspGlnSerGluHisValProValLysLysIle 323  
Qy 946 -----GTCACGCTTGATGCGGCATTCCGAACCGCAAAACGAGACAAGGCGAC-- 993  
Db 324 GlyGluGlnThrAsnHisPheIleSerLysTyrGlyAspLeuHisGluAspLysMetAspMet 343  
Qy 994 -----GCGCCCGCGCTGCGCGTG 1011  
Db 344 LeuThrIleProProArgArgVal 351

RESULT 6  
S77025  
nitrilase (EC 3.5.5.1) - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein s110784  
C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004

C:Accession: S77025

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
o, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S77025

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-346 <KAN>

A:Cross-references: UNIPROT:Q55949; UNIPARC:UPI00000D7133; EMBL:D64005; GB:AB001339; NID

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

C:Keywords: hydrolase

Alignment Scores:

Pred. No.:	1.94e-30	Length:	346
Score:	584.50	Matches:	128
Percent Similarity:	59.5%	Conservative:	53
Best Local Similarity:	42.1%	Mismatches:	114
Query Match:	31.1%	Indels:	9
DB:	2	Gaps:	6

US-09-751-299-1 (1-1041) x S77025 (1-346)

Qy 25 CGCGGCGCGCGGTGTCAGCGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAGAAA 84  
Db 14 ArgAlaAlaAlaGlnIleSerProValLeuPheSerGlnGlnGlyThrMetGluLys 33  
Qy 85 GCGATCGCGCTGATCGAGCAGCGCGCCAGCAGCAGCGTGCCTGATCGCATTCACAGAG 144  
Db 34 ValLeuAspAlaIleAlaAsnAlaAlaLysGlyValGluLeuIleValPheProGlu 53  
Qy 145 ACTTGATCCCGGCTATCCCTTTTGATATGCTGGCGCGCGGCT--TGGGCATG 201  
Db 54 ThrPheValProTyrTyrProTyrPheSerPheValGluProProValLeuMetGlyLys 73

Qy 202 CGCTTCGTCAGCGCTATTTCGAGAA TTCGCTGCGCGCGCAGCAAGCATGTGCAGGCC 261  
Db 74 SerHisLeuLysLeuTyrGlnGlnAlaValThrValProGly---LysValThrGlnAla 92  
Qy 262 CTGCGGATGCGCGCGCGCGCCAGCATGATGTCGTGCGCGGCTATAGCAGCGCGG 321  
Db 93 IleAlaGlnAlaAlaLysThrHisGlyMetValValValLeuGlyValAsnGluArgGlu 112  
Qy 322 GCGCGACCTCTATATAGGCGCAGCATCTTTCGCGCCGATGCGCATCTGATCGCGCG 381  
Db 113 GluGlySerLeuTyrAsnThrGlnLeuIlePheAspAlaAspGlyAlaLeuValLeuLys 132  
Qy 382 CGCCGCAAGCTCAAGCCTACCCATCGCGAGCGCACCGCTGTCGCGGAGAGACGCGACG 441  
Db 133 ArgArgLysIleThrProThrTyrHisGlyArgMetValTyrGlyGlnGlyAspGlyAla 152  
Qy 442 CATCTCGCGGTGCACGATACCGGCATCGGCGCTCGCGCGCTCTGTTGCTGGAGCAC 501  
Db 153 GlyLeuArgThrValAspThrThrValGlyArgLeuGlyAlaLeuAlaCysTrpGluHis 172  
Qy 502 ATCCAGCATTTGTGGAATAGCCATGTACGCCGCGCAGCAAGTCCACGTGCGGTCG 561  
Db 173 TyrAsnProLeuAlaArgTyrAlaLeuMetAlaGlnHisGluGlnIleHisCysGlyGln 192  
Qy 562 TGGCCGAGCTTCAGCCTCTATGCGCGCATGCGCCTATGCGCTCGGACCGAGTCAATACC 621  
Db 193 PheProGlySerMetVal-----GlyGlnIlePheAlaAspGlnMetGluVal----- 208  
Qy 622 GCGCGAAGCCAGATCTACGCGGTGAGGCGCGCTGCTACGTGCGGCTGCGCGGACC 681  
Db 209 ---ThrMetArgHisHisAlaLeuGlnSerGlyCysPheValIleAsnAlaThrGlyTyr 227  
Qy 682 GTTTCGCGCGGATGATCAAGTATTTGTGATACGCCCGCACAAGAGATGTTCTCTCAAG 741  
Db 228 LeuThrAlaGluGln--LysLeuGlnIleThrThrAspGluLysMetHisGlnAlaLeu 246  
Qy 742 GCCGCGCGGTTTGGCATGATTTTCGGCGCGCGCGCGCGCTGCGCGAGCGGCTC 801  
Db 247 SerGlyGlyCysTyrThrAlaIleIleSerProGluGlyLysHisLeuCysGluProIle 266  
Qy 802 CCGGAGACCGAAGAGGACTGCTGTCGCCGATATCGACCTCGCATGATCGCGTTGGCC 861  
Db 267 AlaGlu--GlyGluGlyLeuAlaIleAlaAspLeuAspPheSerLeuIleAlaLysArg 285  
Qy 862 AAGCGCGCGCGCATCCGCGCGCCACTATTACAGCGCGCCGACGTAACGGCGTGTGCTG 921  
Db 286 LysArgMetMetAspSerValGlyHisTyrAlaArgProAspLeuLeuGlnLeuThrLeu 305  
Qy 922 GATCGACGTCCG 933  
Db 306 AsnAsnGlnPro 309

RESULT 7

JQ1613  
cyanide hydratase (EC 4.2.1.66) - imperfect fungus (Gloeocercospora sorghi)

C:Species: Gloeocercospora sorghi

C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004

C:Accession: JQ1613

A:Molecule type: DNA

A:Residues: 1-368 <WAN>

A:Cross-references: UNIPROT:P32964; UNIPARC:UPI0000128C2E; GB:M99044; NID:G168196; PIDN:f

C:Comment: This enzyme converts HCN to formamide.

C:Genetics:

A:Gene: Cht

C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

C:Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase

F,286/Binding site: carbohydrate (Asn) (covalent) #status predicted







Db 73 -----ArgGlyPheArgPheGlyLeuAlaValGlyValHisAsnGluGluGlyArgAsp 90  
QY 208 ---GTCAGCGCTATTTCGAGAAATTCGCTCGTCGGCGGACGACGAGATGGCAGCCCTG 264  
Db 91 GluPheArgAsnTyrHisAlaSerAlaIleLeuValProGlyProGluValGluArgLeu 110  
QY 265 GCGGATCGCGCCCGCCGACGCGCATGCTGTCGGCTATATACGACGCGCGGCG 324  
Db 111 AlaGluLeuAlaGlyLysAsnAsnValHisLeuValMetGlyAlaIleGluLysAspGly 130  
QY 325 GGCAGCCTATATAGGCGCAGCGCATCTTCGCGCCCGATGGCGATGATCGCGCGCGC 384  
Db 131 TyrThrLeuTyrCysThrAlaLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHis 150  
QY 385 CGCAAGCTCAAGCCTTACCCATGCGAGCGCAACGCTGTCGCGGAGGAGACGCGCAT 444  
Db 151 ArgLysValMetProThrSerLeuGluArgCysIleTyrGlyGlnGlyAspGlySerThr 170  
QY 445 CTCGCGGTGACGATACCGCCATCGGGCGCTCGGCGCGCTCTGTGCTGGAGACATC 504  
Db 171 IleProValTyrAspThrProIleGlyLysIleGlyAlaAlaIleCysTyrProGluAsnArg 190  
QY 505 CAGCCATTGTGAAATACGCCATGATCGCCCGCGGACGACAGATCCACGTCCGCTGTG 564  
Db 191 MetProLeuTyrArgThrAlaLeuTyrAlaLysGlyIleGluIleTyrCysAlaProThr 210  
QY 565 CCGAGCTTCAGCCTCTATCGCGGCGCATGCGCTATGCGCTCGGACCGGAGTCAATACCGCC 624  
Db 211 AlaAspTyrSerLeu-----GluTyrGln 218  
QY 625 GCAAGCCAGATCTAC---GCGGTGAGGGCGGCTGTACGTGCTGGCGTCG----- 672  
Db 219 AlaSerMetIleHisIleAlaValGluGlyGlyCysPheValLeuSerAlaHisGlnPhe 238  
QY 673 TGGCGGACCGTTTCG---CCGGAG-----ATGATCAAGGTATTGGTGATACG 717  
Db 239 CysLysArgArgGluPheProGluHisProAspTyrLeuPheAsnAspIleValAspThr 258  
QY 718 CCCGACAGAGATGTTCTCAAGGCGCGCGGCTTTGCCATGATTTTCGGCCCGAC 777  
Db 259 LysGluHisAspProThrValSerGlyGly-----SerValIleIleSerProLeu 276  
QY 778 GGGCGGCGCTGGCGCGGACCGGCTCCCGGAGCGGAAGGAGACTGTGCTGCCGATATC 837  
Db 277 GlyLysValLeuAlaGlyProAsnTyrGluSer--GluGlyLeuValThrAlaAspLeu 295  
QY 838 GACCTCGGCATGATCGCGTTGGCCAAGCGCGCGCGCATCCGCGCGCACTATTTCACGG 897  
Db 296 AspLeuGlyAspIleAlaArgAlaLysLeuTyrPheAspValValGlyHisTyrSerLys 315  
QY 898 CCCGACGTAAACGGCGCTGCTGTGATGACGTCCGGCCCAACGCGTCACGCTTGAT 957  
Db 316 ProAspIlePheAsnLeuThrValAsnGluHisProLysProValThrPheMetThr 335  
QY 958 GCCGCATTGCAACCGCAAAACGAG 981  
Db 336 LysValGluLysAlaGluAspGlu 343  
RESULT 9  
T49147  
nitrilase (EC 3.5.5.1) 1 [imported] - Arabidopsis thaliana  
N;Alternate names: protein T10D17.100  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 31-Dec-2004  
C;Accession: T49147; T52260; T52263  
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25017  
A;Accession: T49147  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-346 <DAN>  
A;Cross-references: UNIPROT:P32961; UNIPARC:UPI0000001AE2; EMBL:AL353865; GSPDB:GN00061;

A;Experimental source: cultivar Columbia; BAC clone T10D17  
R;Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A;Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz)  
A;Reference number: Z26007; MUID:98145459; PMID:9484465  
A;Accession: T52260  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-346 <HIL>  
A;Cross-references: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAA68935.2  
A;Experimental source: cultivar Columbia  
R;Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A;Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidops  
A;Reference number: Z24515; MUID:8022831; PMID:8022831  
A;Accession: T52263  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-346 <BAR>  
A;Cross-references: UNIPARC:UPI0000001AE2; EMBL:U38845; PIDN:AAB05221.1  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: ATSP:T10D17.100; nit1; NIT1  
A;Map position: 3  
A;Introns: 44/1; 104/1; 202/1; 296/1  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase  
Alignment Scores:  
Pred. No.: 8.44e-23 Length: 346  
Score: 467.00 Matches: 121  
Percent Similarity: 48.2% Conservative: 50  
Best Local Similarity: 34.1% Mismatches: 122  
Query Match: 24.8% Indels: 62  
DB: 2 Gaps: 8  
US-09-751-299-1 (1-1041) x T49147 (1-346)  
QY 10 CCCATGACGAAGTATCGCGCGCGCGGTGACGCGCGCGGCTTCCTCGATCTGCAC 69  
Db 21 ProSerThrThrValArgValThrIleValGlnSerSerThrValTyrAsnAspThrPro 40  
QY 70 CGCACAGTCGAGAAAGCGATCGCGCTGATCGACAGCGCGGCAAGCAGACGTGCGCCTG 129  
Db 41 AlaThrIleAspLysAlaGluLysTyrIleValGluAlaAlaSerLysGlyAlaGluLeu 60  
QY 130 ATCGATTCACGAGACTTGATTCGCGCTATCCCTTTGGATATGCTGGCGCGCG 189  
Db 61 ValLeuPheProGluGlyPheIleGlyGlyTyrPro----- 72  
QY 190 GCTTGGGCATGCGCTTC----- 207  
Db 73 ---ArgGlyPheArgPheGlyLeuAlaValGlyValHisAsnGluGluGlyArgAspGlu 91  
QY 208 GTCCAGCGCTATTTCGAGAAATTCGCTCGCGCGGACGACGATGGCAGGCCCTGGCG 267  
Db 92 PheArgLysTyrHisAlaSerAlaIleHisValProGlyProGluValAlaArgLeuAla 111  
QY 268 GATGCGGCGCGCGGACGCGCATGCTGTGCGCGGCTATAGCGAGCGCGGCGCGC 327  
Db 112 AspValAlaArgLysAsnHisValTyrLeuValMetGlyAlaIleGluLysGluGlyTyr 131  
QY 328 AGCCTCTATATAGGCGCAGCGCATCTTCGCGCCCGATGGCGATGATCGCGCGCGCGC 387  
Db 132 ThrLeuTyrCysThrValLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHisArg 151  
QY 388 AAGCTCAAGCCTTACCCATGCGGAGCGCACCGTGTTCGCGGAGGAGACGCGCATCTC 447  
Db 152 LysLeuMetProThrSerLeuGluArgCysIleTyrGlyGlnGlyAspGlySerThrIle 171  
QY 448 GCGGTGACGATACCGCCATCGGGCGCGCTCGCGCGCTGTGCTGGAGACACATCCAG 507  
Db 172 ProValTyrAspThrProIleGlyLysLeuGlyAlaAlaIleCysThrProGluAsnArgMet 191



Db 275 G1yProAsnPheGluSer---GluGlyLeuIleThrAlaAspLeuAspLeuGlyAspVal 293  
QY 853 GCGTTGGCCAGCGCGCGCGCGATCCGGCGGCGCACTATTACAGCGCCGACGTAAACGGG 912  
Db 294 AlaArgAlaIleLeuTyrPheAspSerValGlyHisTyrSerArgProAspValLeuHis 313  
QY 913 CTGCTGCTGATCGACGTCCGGCCCAACCGCTCGTCACGCTTGATGCCGCATTCGAACCG 972  
Db 314 LeuThrValAsnGluHisPro---LysLysProValThrPheIleSerLysValGluLys 332  
QY 973 CAAAACGAGGAC 984  
Db 333 AlaGluAspAsp 336

RESULT 11  
T52259  
nitriase (EC 3.5.5.1) 2 (imported) - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C.Accession: T52259  
R.Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A.Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A.Reference number: Z26007; MUID:98145459; PMID:9484465  
A.Accession: T52259  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-339 <HIL>  
A.Cross-references: UNIPROT:O04907; UNIPARC:UPI00000A5899; EMBL:Y07648; PIDN:CAA68934.3  
A.Experimental source: cultivar Columbia  
C.Genetics:  
A.Gene: nit2  
A.Introns: 37/1; 97/1; 195/1; 289/1  
C.Superfamily: nitriase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:  
Pred. No.: 1.53e-22 Length: 339  
Score: 463.00 Matches: 119  
Percent Similarity: 50.3% Conservative: 54  
Best Local Similarity: 34.6% Mismatches: 127  
Query Match: 24.6% Indels: 44  
DB: 2 Gaps: 8

US-09-751-299-1 (1-1041) x T52259 (1-339)

QY 16 ACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTCGATCTGCACGCACACA 75  
Db 16 ThrIleValArgAlaThrIleValGlnIleSerThrValTyrAsnAspThrProAlaThr 35  
QY 76 GTGAGAAAGCATCGCGCTGATCGACAGCGCGCCAAAGCAGACGTGCGCTGATCGCA 135  
Db 36 LeuGluLysAlaAsnLysPheIleValGlnAlaIleSerLysGlySerGluLeuValVal 55  
QY 136 TTCCCAAGACTTGATTCGGGCTATCCC-----TTTGATATGCTGGCGCGCG 189  
Db 56 PheProGluAlaPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuGlyValGly 75  
QY 190 GCTTGG-----GGCATGCGCTTCGTCAGCGCTATTTCGAGAAATTCGCTGCGCC 240  
Db 76 ValHisAsnGluGluGlyArgAspGluPheArgLysTyrHisAlaIleValIleLysVal 95  
QY 241 GGCAGCAAGCATGGCAGCGCGCTGGCGGATGCGCGCGCGCGCATGCATGCTGTG 300  
Db 96 ProGlyProGluValGluLysLeuAlaGluLeuAlaGlyLysAsnAsnValTyrLeuVal 115  
QY 301 GCCGCTATAGCAGCGCGCGCGCGCTTATATGAGCCAGCGCATCTTCGGCCCC 360  
Db 116 MetGlyAlaIleGlyLysAspGlyTyrThrLeuTyrCysThrAlaLeuPhePheSerPro 135  
QY 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCTACCCATGCGGAGCGCACCGTG 420  
Db 136 GlnGlyGlnPheLeuGlyLysHisArgLysLeuMetProThrSerLeuGluArgCysIle 155

QY 421 TTCGGCAGGAGAGACGCGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 156 TrpGlyGlnGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGly 175  
QY 481 GCGCTCTGTGCTGGAGACACATCCAGCCATGTGCGAATAAGCCCATGTACGCCCGCGAC 540  
Db 176 AlaAlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaLysGly 195  
QY 541 GAACAGATCCACGTCCG-----TCGTGGCCGAGCTTCAGCCTC 579  
Db 196 IleGluLeuTyrCysAlaProThrAlaAspGlySerLysGluTrpGlnSerSerMetLeu 215  
QY 580 TATCGCGGCATGCGCTATGCGCTCGGACCGAGTCAATACCGCGCAAGCCAGATCTAC 639  
Db 216 His-----Ile 217  
QY 640 GCGGTGAGGCGCGCTGCTACGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 699  
Db 218 AlaIleGluGlyGlyCysPheValLeuSerAlaCysGlnPheCys-----LeuArg 234  
QY 700 AAGGTATTGTGATACGCC-----GACAAAGAGATG 732  
Db 235 LysAspPheProAspHisProAspTyrLeuPheThrAspTyrTyrAspAspLysGluPro 254  
QY 733 TTCTCAAGCGCGCGCGGTTTGCCATGATTTTCGGCGCGCGCGCGCGCGCGCGCGCGCG 792  
Db 255 AspSerIleValSerGlnGlySerValIleIleSerProLeuGlyGlnValLeuAla 274  
QY 793 GAGCGCTCCCGGAGACCGAAGAGGAGATGTGTCGCGCATATCGACCTCGCATGATC 852  
Db 275 G1yProAsnPheGluSer---GluGlyLeuIleThrAlaAspLeuAspLeuGlyAspVal 293  
QY 853 GCGTTGCCCAAGCGCGCGCGCGCATCCGGCGCGCACTATTCACGGCGCGCGCGCGCGCG 912  
Db 294 AlaArgAlaLysLeuTyrPheAsp\*\*ValGlyHisTyrSerArgProAspValLeuHis 313  
QY 913 CTGCTGCTGATCGACGTCCGGCCCAACCGCTGTCACGCTTGATCCGCATTCGAACCG 972  
Db 314 LeuThrValAsnGluHisPro---LysLysProValThrPheIleSerLysValGluLys 332  
QY 973 CAAAACGAGGAC 984  
Db 333 AlaGluAspAsp 336

RESULT 12

S22398  
nitriase (EC 3.5.5.1) - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C.Accession: S22398  
R.Bartling, D.; Seedorf, M.; Mithoefer, A.; Weiler, E.W.  
Eur. J. Biochem. 205, 417-424, 1992  
A.Title: Cloning and expression of an Arabidopsis nitriase which can convert indole-3-ac  
A.Reference number: S22398; MUID:92209532; PMID:1555601  
A.Accession: S22398  
A.Molecule type: mRNA  
A.Residues: 1-346 <BAR>  
A.Cross-references: UNIPROT:P32961; UNIPARC:UPI0000130513; EMBL:X63445; NID:gl63399; PIDN:  
C.Superfamily: nitriase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:  
Pred. No.: 2.07e-22 Length: 346  
Score: 461.00 Matches: 120  
Percent Similarity: 48.2% Conservative: 51  
Best Local Similarity: 33.8% Mismatches: 122  
Query Match: 24.5% Indels: 62  
DB: 2 Gaps: 8

US-09-751-299-1 (1-1041) x S22398 (1-346)

QY 10 CCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGCGGTGTTCTCGATCTGCAC 69



Db 21 ProSerThrValAlaArgValThrIleValGlnSerSerThrValTyrAsnAspThrPro 40  
QY 70 CGCACAGTCGAGAAAGCGATCGGCGCTGATCGAGCAGCGCGCCAGACGAGACGTGCGCCTG 129  
Db 41 AlaThrIleAspLysAlaGluLysTyrIleValGluAlaIleSerLysGlyAlaGluLeu 60  
QY 130 ATCGCATCCCAAGACATTGGATCCCGGCTATCCCTTTGGATATGGCTGGCGCGCG 189  
Db 61 ValLeuPheProGluGlyPheIleGlyGlyTyrPro----- 72  
QY 190 GCTTGGGGCATGCGCTTC----- 207  
Db 73 ---ArgGlyPheArgPheGlyLeuAlaValGlyValHisAsnGluGlyArgAspGlu 91  
QY 208 GTCCAGCGCTATTTCGAGAATTGCTCGTCGCGCGCGCAGCAAGCAGTGGCAGCGCCCTGGCG 267  
Db 92 PheArgLysTyrHisAlaSerAlaIleHisValProGlyProGluValAlaArgLeuAla 111  
QY 268 GATGCGGCGCGCGCGCCACGCGCATGCTGTCGCGCGCTATAGCGAGCGCGCGCGCGCG 327  
Db 112 AspValAlaArgLysAsnHisValTyrLeuValMetGlyAlaIleGluLysGluGlyTyr 131  
QY 328 AGCCTCTATATGCGCCAGCGCGATCTTCGCGCGCGCGATGCGATCTGATCGCGCGCGCGCG 387  
Db 132 ThrLeuTyrCysThrValLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHisArg 151  
QY 388 AAGCTCAAGCCTTACCCTATGCGGAGCGCGCGCTTCGCGGAGGAGAGCGGACGCCATCTC 447  
Db 152 LysLeuMetProThrSerLeuGluArgCysIleTrpGlyGlnGlyAspGlySerThrIle 171  
QY 448 GCGGTGACGATACCGCCATCGGCGCGCTTCGCGCGCGCTGTGCTGCTGGAGACACATCCAG 507  
Db 172 ProValTyrAspThrProIleGlyLysLeuGlyAlaAlaIleCysTrpGluAsnArgMet 191  
QY 508 CCATTGTGGAATACGCCATGTACGCGCGCGCGACGACAGATCCACGTCCGCG----- 558  
Db 192 ProLeuTyrArgThrAlaLeuTyrAlaLysGlyIleGluLeuTyrCysAlaIleProThrAla 211  
QY 559 -----TCGTGGCCGAGCTTCAGCCTCTATCGCGCGATGGCCTATGCGCTCGGA 606  
Db 212 AspGlySerLysGluTrpGlnSerSerMetLeuHis----- 223  
QY 607 CCGGAGGTCAATACCGCGCAAGCCAGATCTTACGCGTGCAGGCGCGCTGTACGTGCTG 666  
Db 224 -----IleAlaIleGluGlyGlyCysPheValLeu 233  
QY 667 GCGTCGTGCGCGACCGGTTTCGCGGAGATGATCAAGGTATTGGTGATACGCC----- 720  
Db 234 SerAlaCysGlnPheCysGlnArg-----LysHisPheProAspHisProAspTyr 250  
QY 721 -----GACAAAGAGATGTTCTCTCAAGCGCGCGCGGCGGTTTGGC 759  
Db 251 LeuPheThrAspTrpTyrAspAspLysGluHisAspSerIleValSerGlnGlyGlySer 270  
QY 760 ATGATTTTCGGCGCGCGCGCGCGCTGGCGGAGCGCGTCCCGGAGACCGAAGAGGA 819  
Db 271 ValIleIleSerProLeuGlyGlnValLeuAlaGlyProAsnPheGluSer---GluGly 289  
QY 820 CTGCTGCTGCGCGATATGACCTTCGCGCATGATCGCGTGGCCAAAGCGCGCGCGCATCCG 879  
Db 290 LeuValThrAlaAspIleAspLeuGlyAspIleAlaArgAlaLysLeuTyrPheAspSer 309  
QY 880 GCGGGCACTATTACAGCGCGCGCGAGCTAACGCGGCTGCTGTGATTCGACGTCCGCGCAA 939  
Db 310 ValGlyTyrTyrSerArgProAspValLeuHisLysLeuThrValAsnGluHisPro---Arg 328  
QY 940 CGCGTCGTACGCGCTTGATGCGCGCATTCGAACCGCAAAAGAGGAC 984  
Db 329 LysSerValThrPheValThrLysValGluLysAlaGluAspAsp 343

RESULT 13  
T52262

nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52262  
R:Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A>Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidops  
A/Reference number: 224515; MUID:8022831; PMID:8022831  
A/Accession: T52262  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-339 <BAR>  
A/Cross-references: UNIPROT:P32962; UNIPARC:UPI000016DA4C; EMBL:U3845; PIDN:AAB05220.1  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: NIT2  
A/Map position: 3  
A/Intons: 37/2; 97/1; 195/1; 289/1  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Alignment Scores:  
Pred. No.: 4.38e-22 length: 339  
Score: 456.00 Matches: 118  
Percent Similarity: 50.0% Conservative: 54  
Best Local Similarity: 34.3% Mismatches: 128  
Query Match: 24.2% Indels: 44  
DB: 2 Gaps: 8

US-09-751-299-1 (1-1041) x T52262 (1-339)

QY 16 ACGAAGTATCGCGCGCGCGGTGCAGGCGCGCGGTTCCTCGATCTCGACCGCACA 75  
Db 16 ThrIleValArgAlaThrIleValGlnAlaSerThrValTyrAsnAspThrProAlaThr 35  
QY 76 GTGAGAAAGCGATCGGCGCTGATCGAGCAGGCGCGCAAGCAGACGTGCGCTGATCGCA 135  
Db 36 LeuGlyLysAlaAsnLysPheIleValGluAlaAlaThrLysGlySerGluLeuVal 55  
QY 136 TTCCCAAGACTTGATTCGCGCGCTATCCC-----TTTGGATATGGCTGCGCGCGCG 189  
Db 56 PheProGluAlaPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuGlyValGly 75  
QY 190 GCTTGG-----GCATGCGGCTTCGTCCAGCGCTATTTCGAGAATTCGCTGCGCG 240  
Db 76 ValHisAsnGluGluGlyArgAspGluPheArgLysTyrHisAlaSerAlaIleLysVal 95  
QY 241 GGCACGAAGCAGTGGGAGCGCGCTGGCGGATGCGCGCGCGCGCGCATGCGATGCTG 300  
Db 96 ProGlyProGluValGluLysLeuAlaGluLeuAlaGlyLysAsnAsnValTyrLeuVal 115  
QY 301 GCCGGCTATAGCGGCGCGCGCGCGCGCTCTATATGGCCAGGCGCATCTTCGGCCCC 360  
Db 116 MetGlyAlaIleGluLysAspGlyTyrThrLeuTyrCysThrAlaLeuPhePheSerPro 135  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCACCGAGCGCACCGGTG 420  
Db 136 GlnGlyGlnPheLeuGlyLysHisArgLysLeuMetProThrSerLeuGluArgCysIle 155  
QY 421 TTCGGCGAGGAGACGCGCGCGCGCGCGCAAGCTCAAGCTTTCGAATACCGCATGCGCGCG 480  
Db 156 TrpGlyGlnGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGly 175  
QY 481 GCGCTCTGTTGCTGGGAGACACATCCAGGCTTTCGAATACCGCATGACCGCGCGAC 540  
Db 176 AlaAlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaLysGly 195  
QY 541 GAACAGTCCACGTCGCG-----TCGTGCGCGAGCTTCAGCCTC 579  
Db 196 IleGluLeuTyrCysAlaIleProThrAlaAspGlySerLysGluTrpGlnSerSerMetLeu 215  
QY 580 TATCGCGGATGCGCTATGCGCTCGGACCGGAGGTCAATACCGCGCGCAAGCATCTAC 639  
:::



Db	216	His-----	-----Ile	217
QY	640	GCCTGAGAGCGCTGCTACGTGTGCGCTGCGACCGTTTCCGCGAGATGATC		699
		:	:	
Db	218	AlaIleGluGlyGlyCysPheValLeuSerAlaCysGlnPheCys-----	LeuArg	234
QY	700	AAGTATTGTGATACGCC-----	GACAAGGAGATG	732
Db	235	LysAspPheProAspHisProAspTyrLeuPheThrAspTrpTyrAspAspGluPro		254
QY	733	TTCTCTAAGCGCGCGCGGTTTGCATGATTTTCGGGCGCGCGCGCGCTGGCC		792
Db	255	AspSerIleValSerGlnGlySerValIleIleSerProLeuGlyGlnValIleVala		274
QY	793	GACGCGCTCCCGAGACCGGAAGAGGAGTGTGGTCCCGATATGACCTGGCATGATC		852
Db	275	GlyProAsnPheGluSer--GluGlyLeuIleThrAlaAspLeuAspGlyAspVal		293
QY	853	GCGTTGGCCAAAGCGCGCGCGATCCGCGCGGACCACTATTTCAGCGCGCGCAAGTAACGCGG		912
Db	294	AlaArgAlaLysLeuTyrPheAspSerValGlyHisTyrSerArgProAspValIleuHis		313
QY	913	CTGCTGCTGGATCGACGTCGCGCGCCAAACGCGCTGTACGCTTGATGCGCGCATTCGAACCG		972
Db	314	LeuThrValAsnGlnHisPro--LysLysProValThrPheIleSerLysValGluLys		332
QY	973	CAAAACGAGGAC	984	
		:::		
Db	333	AlaGluAspAsp	336	

## RESULT 14

132266  
nitrilase-like protein [imported] - rice  
C/Species: *Oryza sativa* (rice)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52266  
R/Chiba, R.; Dohmoto, M.; Yamaguchi, K.  
submitted to the EMBL Data Library, May 1999  
A/Description: *Oryza sativa*, a gene for nitrilase-like protein.  
A/Reference number: Z26008  
A/Accession: T52266  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-362 <CHI>  
A/Cross-references: UNIPROT:Q9SXX6; UNIPARC:UPI000009C92A; EMBL:AB027054; PIDN:BAA77679  
A/Experimental source: cultivar Nipponbare  
C/Genetics:  
A/Gene: ONT4  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)

**Alignment Scores:**

Pred. No.:	9.61e-20	Length:	362
Score:	420.00	Matches:	117
Percent Similarity:	50.7%	Conservative:	57
Best Local Similarity:	34.1%	Mismatches:	137
Query Match:	22.3%	Indels:	32
DB:	2	Gaps:	10

US-09-751-299-1 (1-1041) x T52266 (1-362)

```

QY      25  CGCGGCGCGCGCGGTGCAGGCGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAGAA   84
      |||  |||||:::  |||||  |||  |||:::  |||
Db      32  ArgAlaThrValValGlnAlaSerThrValPheTyrAspThrProAlaThrLeuAspLys   51
      |||  |||||:::  |||||  |||  |||:::  |||
QY      85  GCGCATCGGCGCTGATCGAGCAGCGCGCCACAGCAGGACGTGCGCTGATCGATTCGCCAGAG   144
      |||  |||||:::  |||||  |||  |||:::  |||
Db      52  AlaGluArgLeuIleGluGluAlaAlaGlyTyrGlySerGlnLeuValValPheProGlu   71
      |||  |||||:::  |||||  |||  |||:::  |||
QY      145 ACTTGGAATTCGCGGCTATCCCTTTGG-----ATATGGCTG   180
      :::::  |||||  |||  |||  |||  |||
Db      72  AlaPheValGlyGlyTyrProArgGlySerThrPheGlyPheGlyAlaAsnIleSerIle   91
      |||  |||||:::  |||||  |||  |||:::  |||
QY      181 GCGCGCGCGCGCT--TGGGGCATGCGCTTCGTCACGCGCTATTTCGAGATTGCGCTCGTG   237

```

Db	92	GlyAsnProLysAspLysGlyLysGluGluPheArgLysTyrHisAlaAlaAlaIleGlu	111
QY	238	CGCGGCACGACAGCATGTGGCAGGCCCTGGCGGATGCGGCCCGCCGACCGCATGCATGTC	297
Db	112	ValProGlyProGluValThrArgLeuAlaAlaMetAlaGlyLysTyrLysValPheLeu	131
QY	298	GTCGCCGCTATAGCCGAGCGCGCGGGCGGACGCTCTATATATGGCCAGCGCATCTTCGGC	357
Db	132	ValMetGlyValIleGluArgGluGlyTyrThrLeuTyrCysSerValLeuPhePheAsp	151
QY	358	CCCGATGGCGATCTGATCGCGCGCGCGCCGCAAGCTCAAGCCTCAAGCCTCAAGCCTTCG	417
Db	152	ProLeuGlyArgTyrLeuGlyLysHisArgLysLeuMetProThrAlaLeuGluArgIle	171
QY	418	GTCCTTCGGCGAGGAGACGGGACGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTC	477
Db	172	IleTyrGlyPheGlyAspGlySerThrIleProValTyrAspThrProLeuGlyLysIle	191
QY	478	GCGCGGCTCTGTGCTGGAGACATCCAGCCATGTGCGAAATACGCCATGTACGCCGCC	537
Db	192	GlyAlaLeuIleCysTyrGluAsnLysMetProLeuLeuArgThrAlaLeuTyrGlyLys	211
QY	538	GACGAACAGGTCCACGTCGCGTGTGGCGGAGCTTCAGCCTCTATCGCGGCATGGCCTAT	597
Db	212	GlyIleGluIleTyrCysAla-----ProThr-----	220
QY	598	GCGCTCGGACCGGAGGTCATACCGCCGCAAGCCAGATCTACGCGGTCCAGGGCGGCTGC	657
Db	221	AlaAspSerArgGlnValTyrGlnAlaSerMetThrHisIleAlaLeuGluGlyCys	240
QY	658	TACGTGCTGGCGTCG-----TGGCGGACCGTTTCG-----CCGAGATG	696
Db	241	PheValLeuSerAlaAsnGlnPheCysArgArgLysAspTyrProProProGluTyr	260
QY	697	ATCAAGGTA--TTGGTGATACGCCCCGACAAGAGAGATGTTCTCAAGCGCGCGCGGT	753
Db	261	ValPheThrGlyLeuGlyGluGluProSerProAspThrValValCysProGlyGly---	279
QY	754	TTTGCCATGATTTTCGGGGCGCCGCGCGCGCGCTGCGCCGCGCGCTCCCGAGACCGAA	813
Db	280	--SerValIleIleSerProSerGlyGluValLeuAlaGlyProAsnTyrGlu--Gly	297
QY	814	GAGGGACTGTGGTCGCGCATATCGACCTCGGCATGATCGCGTTGGCCAAAGCGGGCGCC	873
Db	298	GluAlaLeuIleThrAlaAspLeuAspLeuGlyGluIleValArgAlaLysPheAspPhe	317
QY	874	GATCCGGCGGCACCTATTTCAGGCGCCGACGTAACGCGGCTGCTGATCGACGTCCG	933
Db	318	AspValValGlyHisTyrAlaArgProGluValLeuSerLeuValValAsnAspGlnPro	337
QY	934	GCCCAACGCGTCTCACCGCTGATCCGCATTCGAACCGCAAAACGAGACAAGGGCGAC	993
Db	338	--HisLeuProValSerPheThrSerAlaAlaGluLysThrThrAlaAlaLysSerAsp	356
QY	994	GCGCCCGCG 1002	
Db	357	SerThrAla 359	

## RESULT 15

103739  
nitrilase (EC 3.5.5.1) 4B - common tobacco  
C/Species: *Nicotiana tabacum* (common tobacco)  
C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C/Accession: T03739  
R;Tsunoda, H.  
submitted to the EMBL Data Library, January 1996  
A/Reference number: Z15044  
A/Accession: T03739  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-348 <TSU>  
A/Cross-references: UNIPROT:Q42966; UNIPARC:UPI000009F332; EMBL:D83078

A;Experimental source: strain SRI, tissue-type leaf  
C;Genetics:  
A;Gene: TNIT4B  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase

Alignment Scores:	
Pred. No.:	3.69e-19
Score:	411.00
Percent Similarity:	52.7%
Best Local Similarity:	35.1%
Query Match:	21.9%
DB:	2
	Gaps: 10
	Length: 348
	Matches: 110
	Conservative: 55
	Mismatches: 118
	Indels: 30
	Gaps: 10

US-09-751-299-1 (1-1041) x T03739 (1-348)

QY	25	CGCGCGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAGAA	84
		:::            :::	
Db	30	ArgAlaThrValValGlnAlaSerThrIlePheTyraSpThrProAlaThrLeuAspLys	49
		:::            :::	
QY	85	GCGATCGCGCTGATCGAGCAGGCGCGCAAGCAGCAGTGCCTGATCGCATTCGCCAG	144
		:::            :::	
Db	50	AlaGluArgLeuLeuAlaGluAlaAlaSerTyrglyAlaGlnLeuValValPheProGlu	69
		:::            :::	
QY	145	ACTTGATTCGCGCTATCCC-----TTTGGATATGCTGGC--GCGCG	189
		:::               :::	
Db	70	AlaPheIleGlyGlyTyrrProArgGlySerThrPheGlyValSerIleGlyAsnArgThr	89
		:::	
QY	190	GCTTGGGCGCATGCGCTTCTGTCAGCGCTATTTCGAGAATTCGCTCGTGC GCGCAGCAG	249
		::::	
Db	90	AlaLysGlyLysGluGluPheArgLysTyrrHisAlaSerAlaIleAspValProGlyPro	109
		::::	
QY	250	CAGTGGCAGCGCCCTGGCGGATGCGCGCGCGCCGACGCGCATGTCATGTCGGCCGCTAT	309
		:::           ::::	
Db	110	GluValAspArgLeuAlaAlaMetAlaGlyLysTyrrLysValTyrrLeuValMetGlyVal	129
		::::	
QY	310	AGCGAGCGCGCGCGCGCGCAGCCTCTATATGSGCCAGCGCATCTTCGCCCCGATGGCGAT	369
		::::	
Db	130	IleGluArgAspGlyTyrrThrLeuTyrrCysThrValLeuPhePheAspSerGlnGlyHis	149
		::::	
QY	370	CTGATCGCGCGCGCGCGCAGCTCAAGCTTACCCTATGCGGAGCGCACCCTGTCGCGGAG	429
		:::       ::::	
Db	150	TyrrLeuGlyLysHisArgLysIleMetProThrAlaLeuGluArgIleIleTyrrGlyPhe	169
		::::	
QY	430	GGAGACGCGCAGCCATCTCGCGGTGACAGTACC GCCATCGGGCGGCTCGCGCTGTGT	489
		:::	
Db	170	GlyAspGlySerThrIleProValTyrrAspThrProLeuGlyLysIleGlyAlaAlaIle	189
		:::	
QY	490	TGCTGGGAGCACATCCAGCCATTGTGAAATACGCCATGTACCGCGCGCAAGAACAGTTC	549
		:::	
Db	190	CysTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyrrAlaLysGlyIleGluIle	209
		:::	
QY	550	CACGTCGCGTCTGTGGCGGAGCTTACGCTCTATCGCGGATGGCCTATGCGCTCGAACCG	609
		:::         :::	
Db	210	TyrrCysAla-----ProThr-----AlaAspSerArg	218
QY	610	GAGGTCAATACCGCGCGCAGCCAGATCTACGCGTGCAGGCGCGGTGCTACTGCTGGCG	669
		:::         :::	
Db	219	AspValTrpGlnAlaSerMetThrHisIleAlaLeuGluGlyGlyCysPheValLeuSer	238
		:::	
QY	670	TCG-----TGCGCGACCGTTTCG-----CCGAGATGATCAAGTATTG	708
		:::	
Db	239	AlaAsnGlnPheCysArgArgLysAspTyrrProProProGluTyrrValPheSerGly	258
		:::	
QY	709	GTGAT-----ACGCCGACAGAGAGATGTTCTCAAGCGCGCGCGGCTTTTGCCATG	762
		:::         :::	
Db	259	ThrGluAspLeuThrProAspSer-----IleValCysAlaGlyGly-----SerVal	274
		:::	
QY	763	ATTTTCGGGCGCGCGCGCGCTGCGCGAGCGCTCCCGAGACCGAAGAGGAGGACTG	822
Db	275	IleIleSerProSerGlyAlaValLeuAlaGlyProAsnTyrrGlu---GlyGluAlaLeu	293
QY	823	CTGTGCGCGGATATCGACTTCGCGCATGATCGCGTTGGCCAGGCGCGCGCGGATCCGGCG	882

Db 294 I l e S e r A l a A s p l e u A s p l e u G l u I l e A l a A r g A l a l a l y s p h e a s p h e a s p v a l v a l 313  
::: |||||:::||||| ||||| ||||| |||||

QY 883 G G C C A C T A T T C A C G G C C C G A C G T A A C C G G C T G C T G 921  
|||||:::||||| |||||::: |||||

Db 314 G l y h i s t y r A l a A r g p r o g l u v a l l e u s e r l e u l e v a l 326  
|||||:::||||| |||||::: |||||

Search completed: April 27, 2006, 01:38:18  
Job time : 48.019 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:07:06 ; Search time 51.2142 Seconds  
(without alignments)  
2868.171 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1881

Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgacgtag 1041

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 segs, 705528306 residues

Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090034\_18631/app\_query.fasta\_1  
-DB=UniProt -QFMT=faatan -SUFFIX=rup -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US09751299\_@CCN\_1\_1\_580\_@runat\_26042006\_090034\_18631 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80:\*

1: uniprot\_sprot:\*\n2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1806	96.0	346	2	Q6RWK5_9ZZZZ	Q6rwk5 uncultured
2	1464	77.8	341	2	Q6RWR4_9ZZZZ	Q6rwr4 uncultured
3	1209	64.3	348	2	Q6RWS2_9ZZZZ	Q6rws2 uncultured
4	1165	61.9	353	2	Q6RWF9_9ZZZZ	Q6rwf9 uncultured
5	1155	61.4	353	2	Q6RWF0_9ZZZZ	Q6rwf0 uncultured
6	1132.5	60.2	358	2	Q6RWI0_9ZZZZ	Q6rwi0 uncultured
7	1130	60.1	354	2	Q6RWK6_9ZZZZ	Q6rwk6 uncultured
8	1127	59.9	331	2	Q706Q8_PSEPU	Q706q8 pseudomonas
9	1125	59.8	345	2	Q6RWS5_9ZZZZ	Q6rws5 uncultured
10	1099.5	58.5	334	2	Q6RWE3_9ZZZZ	Q6rwe3 uncultured
11	1076.5	57.2	337	2	Q6RWN9_9ZZZZ	Q6rwn9 uncultured
12	1073.5	57.1	338	2	Q6RWG6_9ZZZZ	Q6rwg6 uncultured
13	1070	56.9	336	2	Q6RWF6_9ZZZZ	Q6rwf6 uncultured
14	1068.5	56.8	335	2	Q6RWF7_9ZZZZ	Q6rwf7 uncultured
15	1068.5	56.8	336	2	Q5YUM5_NOCFA	Q5yum5 nocardia fa
16	1066.5	56.7	337	2	Q6RWN8_9ZZZZ	Q6rwn8 uncultured

17	1064.5	56.6	338	2	Q6RWN7_9ZZZZ	Q6rwn7 uncultured
18	1064.5	56.6	350	2	Q5EG61_PSEFL	Q5eg61 pseudomonas
19	1057.5	56.2	352	2	Q6RWH0_9ZZZZ	Q6rwh0 uncultured
20	1052.5	56.0	338	2	Q6RWE4_9ZZZZ	Q6rwe4 uncultured
21	1044.5	55.5	337	2	Q6RWS3_9ZZZZ	Q6rws3 uncultured
22	1043.5	55.5	338	2	Q6RWM2_9ZZZZ	Q6rwm2 uncultured
23	1041	55.3	336	2	Q6RWJ5_9ZZZZ	Q6rwj5 uncultured
24	1038.5	55.2	338	2	Q6RWQ4_9ZZZZ	Q6rwq4 uncultured
25	1038.5	55.2	338	2	Q6RWR6_9ZZZZ	Q6rwr6 uncultured
26	1028.5	54.7	337	2	Q6RWI7_9ZZZZ	Q6rwi7 uncultured
27	1026.5	54.6	338	2	Q6RWH1_9ZZZZ	Q6rwh1 uncultured
28	1025.5	54.5	336	2	Q500U1_PSESY	Q500u1 pseudomonas
29	1023	54.4	337	2	Q6RWQ3_9ZZZZ	Q6rwq3 uncultured
30	1023	54.4	337	2	Q6RWQ7_9ZZZZ	Q6rwq7 uncultured
31	1020.5	54.3	327	2	Q6RWH9_9ZZZZ	Q6rwh9 uncultured
32	1012	53.8	347	2	Q88B32_PSESM	Q88b32 pseudomonas
33	1008	53.6	349	2	Q6RWS0_9ZZZZ	Q6rws0 uncultured
34	1007	53.5	337	2	Q6RME0_9ZZZZ	Q6rme0 uncultured
35	1004.5	53.4	337	2	Q6RWM3_9ZZZZ	Q6rwm3 uncultured
36	986	52.4	345	2	Q6RWN2_9ZZZZ	Q6rwn2 uncultured
37	974.5	51.8	337	2	Q6RWM9_9ZZZZ	Q6rwm9 uncultured
38	972	51.7	334	2	Q6RWC9_9ZZZZ	Q6rwc9 uncultured
39	964	51.2	334	2	Q89GE3_BRAJA	Q89ge3 bradyrhizob
40	949.5	50.5	346	2	Q6RWM1_9ZZZZ	Q6rwm1 uncultured
41	949	50.5	333	2	Q6RWL0_9ZZZZ	Q6rwl0 uncultured
42	945	50.2	332	2	Q6RWJ0_9ZZZZ	Q6rwj0 uncultured
43	869	46.2	356	1	NRLA_ALCFA	P20960 alcaligenes
44	829.5	44.1	365	1	NRL2_RHORH	Q03217 rhodococcus
45	825.5	43.9	383	1	NRL1_RHORH	Q02068 rhodococcus

ALIGNMENTS

RESULT 1	Q6RWK5_9ZZZZ	PRELIMINARY;	PRT;	346 AA.
ID	Q6RWK5_9ZZZZ			
AC	Q6RWK5;			
DT	05-JUL-2004 (Tremblrel. 27, Created)			
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)			
DT	05-JUL-2004 (Tremblrel. 27, Last annotation update)			
DE	Nitrilase (EC 3.5.5.7).			
GN	ORFNames=BD5086;			
OS	uncultured organism.			
OC	unclassified; environmental samples.			
OX	NCBI_TaxID=155900;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RX	PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;			
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M., Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K., McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M., Mathur E., Kretz P.L., Burk M.J., Short J.M.;			
RT	"Exploring nitrilase sequence space for enantioselective catalysis.";			
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).			
DR	EMBL; AY487497; AAR97444.1; -; Genomic DNA.			
DR	GO; GO:0018762; F:aliphatic nitrilase activity; IEA.			
DR	GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.			
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.			
DR	InterPro; IPR000132; N/CN_hydrtase.			
DR	InterPro; IPR003010; NtIse/CNhydrtse.			
DR	Pfam; PF00795; CN_hydrolase; 1.			
DR	PROSITE; PS00263; CN_HYDROLASE; 1.			
DR	PROSITE; PS00921; NITRIL_CHT_2; 1.			
KW	Hydrolase.			
SQ	SEQUENCE 346 AA; 37432 MW; 12FE01C35AD463CE CRC64;			
Alignment Scores:				
Pred. No.:	4.47e-100	Length:	346	
Score:	1806.00	Matches:	346	
Percent Similarity:	100.0%	Conservative:	0	
Best Local Similarity:	100.0%	Mismatches:	0	
Query Match:	96.0%	Indels:	0	
DB:	2	Gaps:	0	

US-09-751-299-1 (1-1041) x Q6RWR5\_9ZZZZ (1-346)

QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGTGCAGGCCCGCGGTGTTCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaIaIaProValPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCCAAAGCAGGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaIaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCCTCCAGAGACTTGATCCCGGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTripleProGlyTyrProPheTripleTripleu 60  
QY 181 GCGCGCGCGCTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAATTGCTGTCGC 240  
Db 61 GlyAlaProAlaTripleGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTCGACAGGCCCTGCGCGGATGCGGCCCGCCAGCGCATGTCGTG 300  
Db 81 GlySerLysGlnTripleGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCAGCGCGCGCGCGCGCGCTCTATATGGGCCAGCGCATCTTGGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGAGCTCAAGCCTAACCCATGCGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGGAGCACATCCAGCCATTGTCGAATACGCCATGTACGCGCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCAGCTCGCGCTCGTGGCGCGAGCTTCAGCCTCTATCGCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaIaSerTrpSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGCAGGGCGGTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyGlyCysTyr 220  
QY 661 GTGCTGGCGGTGTCGCGCAGCCGTTTCGCGGAGATGATCAAGGTATGGTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAAGGAGATGTTCTCAAGGCGCGCGGGGCTTTGCCATGATTTTCGGGCGCCAGCGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCGCTGGCGCAGCGCGCTCCCGAGACCGAAGAGGAGACTGCTGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluLysLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCGCATCCGCGCGGCCCACTATTACGGGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGGCGGTGCTGCTGATGACGTCGCGGCCCAACGCGTGTGCAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGGACGAGGCGCAGCGCGCGCGCTGCGCGGTGTCGGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValaGln 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
|||||

Db 341 SerAlaAlaAlaAlaGln 346  
RESULT 2  
Q6RWR4\_9ZZZZ PRELIMINARY; PRT; 341 AA.  
AC Q6RWR4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7266;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis.",  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487438; AAR97385.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; Ntase/CNhydrtase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 341 AA; 36918 MW; 23AC5387F4C1FDEE CRC64;  
  
Alignment Scores:  
Pred. No.: 1.44e-79 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 2 Gaps: 0  
  
US-09-751-299-1 (1-1041) x Q6RWR4\_9ZZZZ (1-341)  
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGTGCAGGCCCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaIaIaProSerPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCCAAAGCAGGAC 120  
Db 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaIaGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCCTCCAGAGACTTGATCCCGGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTripleProGlyTyrProLeuTripleTripleu 60  
QY 181 GCGCGCGCGCTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAATTGCTGTCGC 240  
Db 61 GlySerProAlaTripleGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTCGACAGGCCCTGCGCGGATGCGGCCCGCCAGCGCATGTCGTG 300  
Db 81 GlySerLysGlnTripleAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCGGCTATAGCAGCGCGCGCGCGCGCGCTCTATATGGGCCAGCGCATCTTGGCCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGCAGCTCAAGCCTAACCCATGCGGAGCGCACCGTG 420  
Db 121 GluGlyLysLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140



QY 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGTGACGATACCGCCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTGCTGGAGACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCCGAC 540  
Db 161 AlaLeuCyAsCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAsn 180  
QY 541 GAACAGGTCCACGTCGCGCTGCTGGCCGAGCTTCACTCTATCGCGGATGGCCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCyAsPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCGGAGATGATCAAGGTATTGTGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGGCGCGAGCGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCTGCCATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTGGCATGATCGCTTGCCCAAGCGCGCGCGGATCCGGCGGGCCACTATTTCAGCGGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTAACGCGGCTGCTGTCGATCGACGTCGCGGCCCAAGCGCTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGCAACCGCAAAAGAGACAAAGGGCGAGCGGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
RESULT 3  
Q6RWS2\_9ZZZZ PRELIMINARY; PRT; 348 AA.  
ID Q6RWS2\_9ZZZZ AC Q6RWS2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE NitriIase (EC 3.5.5.7).  
GN ORFNames=BD5338;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriIase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487430; AAR97377.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitriIase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; N:ise/CNhydase.  
DR InterPro; IPR002332; PII\_GlnB\_UMP\_S.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN 1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN 1.

KW Hydrolase.  
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;  
Alignment Scores:  
Pred. No.: 2,83e-64 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 2 Gaps: 2  
US-09-751-299-1 (1-1041) x Q6RWS2\_9ZZZZ (1-348)  
QY 10 CCCATGACGAAG--TATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCGAAGAAGCGATCGGCTGATCGACAGCGCGCGCAACAGACGTGGC 126  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaIaSerAsnGlyAlaLys 41  
QY 127 CTGATCGCAATTCCAGAGACTTGATTTCCGGCTATCCCTTTGGATATGCTGGCGCG 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGCGATGCGCTTCTGTCACGCTATTTCGAGAATTGCTGTCGCGCGACG 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGGCCCTGCGGATGCGGCGCGCGCGCAAGCATGTCGTGGCGCGC 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnLysIleTyrValValMetGly 101  
QY 307 TATACGACGCGCGCGCGCGCGCACTTATATGGCCAGCGCATTTCCGCCCGATGCG 366  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCATCGGAGCGACCGTGTGCGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGCGACGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCysValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTTCGTGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCAGCAACAG 546  
Db 162 CysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGln 181  
QY 547 GTCCACGTGCGCTGCTGGCCGAGCTTCACTCTATCGCGCGCATGSCCTATGCGCTCGA 606  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTCTACGTGCTG 666  
Db 202 ProGluLeuAsnAsnAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTGTGCGCGACCGTTTCGCGCGGAGATGATCAAGGTATTGTGATACGCCCGACAAG 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGCGCGCGCGGCTTTTGCCATGATTTTCGGGCGCGAGCGCGGCC 786  
Db 242 GluProLeuLeuLeuGluGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGGCGAGCCGCTCCCGGAGACCGAAGAGGAGACTGCTGCTGCGCATATCGACTCGGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGCGGCCACTATTTCAGCGCGCGAGTA 906  
Db 282 MetIleSerMetAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301

QY 907 ACGCGGCTGCTGATCGACGTCGCCCAACGCGTGTACGCTTGATGCCGATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGACAGAGGGCGACGGCCCGCGCTGCGGTGTGGCGAAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGlnGln 340  
QY 1027 GCCCGCCGCG 1035  
Db 341 LysAlaAla 343  
RESULT 4  
Q6RMWF9\_9ZZZZ PRELIMINARY; PRT; 353 AA.  
ID Q6RMWF9\_9ZZZZ  
AC Q6RMWF9;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrlase (EC 3.5.5.7).  
GN ORFNames=BD7753;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrlase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487543; AAR97490.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrlase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydrtse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KM Hydrolase.  
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC31326 CRC64;  
Alignment Scores:  
Pred. No.: 1.23e-61 Length: 353  
Score: 1165.00 Matches: 229  
Percent Similarity: 78.8% Conservative: 31  
Best Local Similarity: 69.4% Mismatches: 66  
Query Match: 61.9% Indels: 4  
DB: 2 Gaps: 3  
US-09-751-299-1 (1-1041) x Q6RMWF9\_9ZZZZ (1-353)  
QY 19 AAGTATCGCGCGCGCGGTGACAGCGCGCGGTTCCTCGATCTGACCGCACAGTC 78  
Db 7 LysTyrLysValAlaAlaValGlnAlaAlaProAlaPheLeuAspLeuAspAlaSerVal 26  
QY 79 GAGAAAGCATCGGCTGTATCGAGCAGGCGGCCAAGCAGGACGTGGCCTGATGCATTTC 138  
Db 27 GluLysAlaValArgPheIleAspGluAlaGlyAlaAlaGlyAlaArgLeuIleAlaPhe 46  
QY 139 CCAGAGACTTGATATCCCGGCTATCCCTTTTGGATATGGCTGGGCGCGCGCTTGGGGC 198  
Db 47 ProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuGlyAlaProAlaTrpAla 66  
QY 199 ---ATGCGC---TTCGTCCAGCGCTATTTCGAGAATTGGCTCGGCGCGCGCAGCAAGCAG 252  
Db 67 IleMetArgGlyPheValSerArgTyrPheAspAsnSerLeuSerTyrAspSerProGln 86  
QY 253 TGGCAGGCGCTGCGGATGCGGCGCGCGCGCCACGCGATGATGTGTGCGCGCTATAGC 312

Db 87 AlaGluLysLeuArgAlaAlaAlaLysArgAsnLysMetValValLeuGlyLeuSer 106  
QY 313 GAGCGCGCGCGCGCAGCCTCTATATGGGCCAGCGCATCTTGGCGCCGATGGCGATCTG 372  
Db 107 GluArgAspGlyGlySerLeuTyrIleAlaGlnTrpIleIleGlyProAspGlyGluThr 126  
QY 373 ATGCGCGCGCGCGCGCAAGCTCAAGCCTAACCCATGCGGAGCGCAGCGTGTGGCGAGGA 432  
Db 127 IleAlaLysArgArgLysLeuLysProThrHisAlaGluArgThrValPheGlyGluGly 146  
QY 433 GACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGCGCGCTGCGCGCTGTTCG 492  
Db 147 AspGlySerHisLeuAlaValHisGluLeuAspValGlyArgLeuGlyAlaLeuCysCys 166  
QY 493 TGGGAGCATCCAGCCATTGTGAAATACGCCATGTACGCCCGCAGCAACAGTTCAC 552  
Db 167 TrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHis 186  
QY 553 GTGCGGTGTCGCGCGAGCTTCAGCCTTATCGCGCATGCGCCTATGCGCTCGACCGAG 612  
Db 187 ValAlaAlaTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGlu 206  
QY 613 GTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTCTACGTGCGGTGC 672  
Db 207 ValAsnAsnAlaAlaSerLysIleTyrAlaValGluGlySerCysPheValIleAlaPro 226  
QY 673 TGGCGCAGCGTTTCGCGCGAGATGATCAAGTATTGGTGATACGCCCGACAAAGAGATG 732  
Db 227 CysAlaThrValSerGlnAlaMetIleAspGluLeuCysAspThrProGluLysHisGln 246  
QY 733 TTCTCAAGCGCGCGCGGCTTTTGCCATGATTTTCGGCGCGCGCGCGCGCTGTGCC 792  
Db 247 PheLeuHisAlaGlyGlyPheAlaValIleTyrGlyProAspGlyAlaProLeuAla 266  
QY 793 GAGCGCTCCCGAGAGACCGAGAGGAGGACTGTGTGCGCGCATATCGACCTCGGATGATC 852  
Db 267 AlaProLeuProProAspLysGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIle 286  
QY 853 GCGTTGGCCAAGCGCGCGCGCATCCGCGGCGCACTATTCAAGCGCGCGCGCATGACGG 912  
Db 287 SerValAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspValThrArg 306  
QY 913 CTGCTGTGATCGACGTCGCGGCCCAACGCGTGTGTACGCTTGATGCCGATTCGAACCG 972  
Db 307 LeuLeuPheAsnAsnArgProGlyTyrArgValGluThrMet-----AlaLeuProIle 324  
QY 973 CAAAACGAGACAGAGGCGCGCGCGCGCGCG 1002  
Db 325 AspAlaGluThrLysAlaGluAlaProAla 334  
RESULT 5  
Q6RMWF0\_9ZZZZ PRELIMINARY; PRT; 353 AA.  
ID Q6RMWF0\_9ZZZZ  
AC Q6RMWF0;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrlase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrlase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487552; AAR97499.1; -; Genomic\_DNA.





QY 199 ---ATGCGC---TTGCTCCAGCGCTATTTCGAAATTCGCTCGTGCGCGCAGCAAGCAG 252  
Db 67 IleMetArgGlyPheValSerArgTyrPheAspAsnSerLeuSerTyrAspSerArgIn 86  
QY 253 TGGCAGCCCTGGCGGATGCGCGCCGCGCCAGCATGCATGTCTGGCGGCTATAGC 312  
Db 87 AlaGluArgLeuArgAspAlaAlaLysArgHisLysLeuThrValValMetGlyLeuSer 106  
QY 313 GAGCGCGCGCGCGCAGCCTCTATATGGCCAGCGCATCTTCGCGCCCGATGGCGATCTG 372  
Db 107 GluArgAlaGlyGlySerLeuTyrIleAlaGlnTrpIleIleGlyProAsnGlyGluThr 126  
QY 373 ATCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCCTGTTCCGCGAGGA 432  
Db 127 ValAlaGlnArgArgLysLeuLysProThrHisAlaGluArgThrValPheGlyGluGly 146  
QY 433 GACGCGACCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGCGCGCTCTGTGC 492  
Db 147 AspGlySerHisLeuAlaValHisAsnLeuProIleGlyArgLeuGlyAlaLeuCysCys 166  
QY 493 TGGAGCACATCCAGCCATTGTGCAATACGCCATGTACGCGCCGCAAGCAGAGTCCAC 552  
Db 167 TrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGluHis 186  
QY 553 GTCGCGTGTGGCGCGAGCTTCAGCCTCTATCGCGCGCATGGCCTTATGCGTCCGACCGAG 612  
Db 187 ValAlaAlaIleTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGlu 206  
QY 613 GTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTACGTGCTGGCGTGC 672  
Db 207 ValAsnAsnAlaAlaSerGlnIleTyrAlaValGluGlySerCysPheValAlaPro 226  
QY 673 TGGCGGACGTTTTCGCGGAGATGATCAAGTATTGTTGATACGCGCGCAAGAGAGATG 732  
Db 227 CysAlaValIleSerGlnGluMetIleAspLeuMetCysAspThrProAspLysHisGln 246  
QY 733 TTCTTCAAGCGCGCGCGGCTTTTTCGATGATTTTTCGCGCGCGCGCGCGCTGGCC 792  
Db 247 LeuIleHisValGlyGlyPheThrValIleTyrGlyProAspGlyAlaArgIleGly 266  
QY 793 GAGCGCGTCCCGGAGAGACCGAAGAGGAGTCTGTCGCCGATATCGACCTCGCATGATC 852  
Db 267 AspLysLeuAlaProAspGlnGluGlyIleValTyrAlaAspIleAspLeuGlyMetIle 286  
QY 853 GCGTTGGCGCAAGCGCGCGCGCGCATCCGCGCGCGCCACTATTCACGCGCCGACGTAACGCG 912  
Db 287 ProIleAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspValThrArg 306  
QY 913 CTGCTGTGATCGAGCTCCGCGCCCAACGCGTCTCACGCTT----- 954  
Db 307 LeuLeuPheAsnAsnArgProAlaAsnArgValGluThrLeuValLeuProValAspGln 326  
QY 955 -----GATCGCGCATTCGAA-----CCGCAA----- 975  
Db 327 ValArgAspIleAspAlaArgValGluAlaAlaProGlnAlaArgProAlaThrGly 346  
QY 976 AACGAGACAAAGGCGCAGCGCCC 999  
Db 347 AsnGluAspProAlaAlaLysPro 354

RESULT 7  
Q6RWK6\_9Z2ZZZ  
ID Q6RWK6\_9Z2ZZZ PRELIMINARY; PRT; 354 AA.  
AC Q6RWK6;  
DT 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)  
DE NitriIase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]

RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriIase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487496; AAR97443.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitriIase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR00132; N/CN hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydase.  
DR pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 354 AA; 37879 MW; 8319C293775ABE923 CRC64;

Alignment Scores:  
Pred. No.: 1.55e-59 Length: 354  
Score: 1130.00 Matches: 224  
Percent Similarity: 76.4% Conservative: 42  
Best Local Similarity: 64.4% Mismatches: 74  
Query Match: 60.1% Indels: 8  
DB: 2 Gaps: 4

US-09-751-299-1 (1-1041) x Q6RWK6\_9Z2ZZZ (1-354)

QY 19 AAGTATCGCGCGCGCGGTGCAAGCGCGCGGCTTCTCGATCTGCACCGCAGACTC 78  
Db 7 LysPheLysValAlaAlaValGlnAlaAlaProAlaPheLeuAspLeuAspAlaSerVal 26  
QY 79 GAGAAAGCATCGGCTTGATCGAGCAGCGCGCCAAAGCAGACGTGCGCTGATCGATTTC 138  
Db 27 GluLysAlaValArgLeuIleAspGluAlaGlyAlaAlaGlyAlaArgLeuIleAlaPhe 46  
QY 139 CCAGAGACTTGATTCGCGGTATCCCTTTGATATGCTGGCGCGCGCTTGGGCGC 198  
Db 47 ProGluThrPheIleProGlyTyrProTrpTrpIleTrpLeuGlyAlaProAlaTrpAla 66  
QY 199 ---ATGCGC---TTGCTCCAGCGCTATTTCGAGAATTCGCTCGTGCGCGCAGCAGCAG 252  
Db 67 IleMetArgGlyPheValSerArgTyrPheAspAsnSerLeuGlnTyrGlyThrProGlu 86  
QY 253 TGGCAGGCCCTGGCGGATGCGCGCCGCGCCAGCGCATGTCGTGCGCGGCTATAGC 312  
Db 87 AlaAspArgLeuArgAlaAlaAlaLysArgAsnLysMetPheValAlaLeuGlyLeuSer 106  
QY 313 GAGCGCGCGCGCGCAGCCTCTATATGGCCAGCGCATCTTCGCGCCCGATGGCGATCTG 372  
Db 107 GluArgAspGlyGlySerLeuTyrIleAlaGlnTrpIleIleGlyProAspGlyGluThr 126  
QY 373 ATCGCGCGCGCGCGCAAGCTCAAGCCTTACCCATGCGGAGCGCACCCTGTTCCGCGAGGA 432  
Db 127 ValAlaThrArgArgLysLeuLysProThrHisAlaGluArgThrValPheGlyGluGly 146  
QY 433 GACGCGACCATCTCGCGGTGCACGATACCGGCATCGGGCGCCTCGCGCGCTGTGTTGC 492  
Db 147 AspGlySerHisLeuAlaValHisGluLeuAspIleGlyArgValGlyAlaLeuCysCys 166  
QY 493 TGGAGCACATCCAGCCATTGTGCAATACGCCATGTACGCCGCGCAAGCAGAGTCCAC 552  
Db 167 TrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGluHis 186  
QY 553 GTCGCGTGTGGCGCGAGCTTCAGCCTCTATCGCGCGCATGGCCTTATGCGTCCGACCGAG 612  
Db 187 IleAlaAlaIleTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGlu 206  
QY 613 GTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTACGTGCTGGCGTGC 672



```
Db 207 ValAsnAsnAlaIaSerLysIleTyrAlaValGluGlySerCysPheValIleAlaPro 226
QY 673 TGGCGGACCGTTTCGCCGAGATGATCAAGGTAATTGGTATACGCCCGCAAGAGATG 732
    |||||
Db 227 CysAlaThrValSerGlnAlaMetIleAspGluLeuCysAspSerProGluLysHisGln 246
QY 733 TTCCTCAAGCGCGCGCGGTTTTCCTATGATTTCCGGCCCGACGCGCGCGCTGGCC 792
    |||||
Db 247 PheLeuHisValGlyGlyPheAlaValIleTyrGlyProAspGlyAlaProLeuAla 266
QY 793 GAGCCGCTCCCGGAGACCGAAGAGGAGCTGCTGTCGCCGATATGACCTCGCATGATC 852
    ::::|
Db 267 LysProLeuAlaProAspGlnGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIle 286
QY 853 GCGTTGGCCAAGCGCGCGCGATCCGCGCGGCACTATTCAAGCGCGCGCACTAACGCGG 912
    ::::|
Db 287 SerValAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspValThrArg 306
QY 913 CTGCTGCTGATCGACGTCCTCGGCCCAACGCGCTGTCACGCTTGATGCCGCAATCGAACCG 972
    |||||
Db 307 LeuLeuPheAsnAsnArgProGlyAsnArgValGluThrLeuAlaLeuProValaAspGln 326
QY 973 CAAACGAG--GACAAAGGCGGACGCGCGCGCGCTGCGCGCTGGCGGAAGCGCC-- 1026
    ::::|
Db 327 GluAlaGluAlaGlyAlaGlyGlyLysProAlaProLysSerProSerValaAlaAlaPhe 346
QY 1027 -----GCCGCGCGCGCAG 1038
    |||||
Db 347 ThrLeuThrGlnAlaAlaAlaGlu 354

RESULT 8
Q706Q8_PSEPU
ID Q706Q8_PSEPU PRELIMINARY; PRT; 331 AA.
AC Q706Q8;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Predicted amidohydrolase/nitrilase.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RX MEDLINE=99008981; PubMed=9791097;
RA Ravatn R., Studer S., Zender A.J.B., van der Meer J.R.;
RT "Int-B13, an unusual site-specific recombinase of the bacteriophage P4
RT integrase family is responsible for chromosomal insertion of the 105-
RT kb c1c-element of Pseudomonas sp. strain B13.";
RL J. Bacteriol. 180:5505-5514(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RX MEDLINE=22708816; PubMed=12823813;
RX DOI=10.1046/j.1365-2958.2003.03548.x;
RA Sentschilo V., Zehnder A.J.B., van der Meer J.R.;
RT "Characterization of two alternative promoters and a transcription
RT regulator for integrase expression in the c1c catabolic genomic island
RT of Pseudomonas sp. strain B13.";
RL Mol. Microbiol. 49:93-104(2003).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RA Gaillard M., Werlen C., Vallaes T., Vorhoeelter F.J., Puehler A.,
RA van der Meer J.R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ617740; CAB92892.1; -; Genomic DNA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN_hydrtase.
DR InterPro; IPR003010; Ntlse/CNhydrtse.
DR Pfam; PF00795; CN_hydrolase; 1.
```

```
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 331 AA; 36554 MM; DEF59D03874D9187 CRC64;

Alignment Scores:
Pred. No.: 2.33e-59 Length: 331
Score: 1127.00 Matches: 206
Percent Similarity: 79.4% Conservative: 45
Best Local Similarity: 65.2% Mismatches: 65
Query Match: 59.9% Indels: 0
DB: 2 Gaps: 0

US-09-751-299-1 (1-1041) x Q706Q8_PSEPU (1-331)
QY 13 ATGACGAAGTATCGCGCGCGCGGTGTCAGCGCGCGGTGTTCTGATCTGACCGC 72
    ::|
Db 6 IleAlaLysTyrLysValAlaAlaValGlnAlaAlaProGluPheLeuAsnLeuAspLys 25
QY 73 ACAGTCGAGAAAGCATCGCGCTGATCGAGCAGCGCGCAAGAGACGTCGCGCTGATC 132
    |||||
Db 26 GlyValGluLysAlaValaValArgLeuIleGluGluAlaAlaLysAsnGlyAlaLysLeuIle 45
QY 133 GCATTCCCAAGACTTGATTCGCGCTATCCCTTTGGATATGCTGGCGCGCGCT 192
    |||||
Db 46 AlaPheProGluValThrLeuProGlyTyrProThrTriPleThrLeuAspSerProAla 65
QY 193 TGGGGCATGCGCTTCGTCACGCGCTATTTCGAGAATTGCTGCTGCGCGCGCAAGACAG 252
    |||||
Db 66 TrpGlyMetGlnPheValGlnArgHisPheGluAsnAlaLeuLeuValGlySerProGln 85
QY 253 TGGCAGGCGCTGAGCGATGCGCGCGCGCGCAAGCATGATGCTGCGCGCTATAGC 312
    ||::|
Db 86 TrpGluArgLeuCysAlaAlaAlaAlaGluHisArgIlePheValValLeuGlyPheCys 105
QY 313 GAGCGCGCGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGCGCGCGATGCGATCTG 372
    |||||
Db 106 GluArgGlnAspGlyThrLeuTyrIleAlaGlnAlaIleIleAspAspGluGlyArgVal 125
QY 373 ATCGCGCGCGCGCGCAAGCTCAAGCCTTACCCATGCGGAGCGCACCGCTGTCGGCAGGGA 432
    ::::|
Db 126 ValSerThrArgArgLysLeuLysProThrHisAlaGluArgThrValTyrGlyGluGly 145
QY 433 GACGCGAGCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGSGCGCTGTGTC 492
    |||||
Db 146 AspGlySerHisLeuSerValHisGlnThrSerIleGlyArgIleGlyAlaLeuSerCys 165
QY 493 TGGAGACATCCAGCCATTTGTGAAATACGCCATGTACGCGCGCGCAAGACAGTCCAC 552
    |||||
Db 166 AlaGluHisIleGlnProLeuSerLysTyrAlaMetTyrSerGlnAsnGluGlnIleHis 185
QY 553 GTCGCGTGGNGCGCGAGCTTCAAGCCTTATCGCGGCGCATGCGCTATGCGCTCGACCGGAG 612
    |||||
Db 186 ValAlaAlaTrpProSerPheSerValTyrArgGlyAlaAlaPheGlnLeuSerProGlu 205
QY 613 GTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTACGCTGCGCTCG 672
    |||||
Db 206 AlaAsnLeuAlaIaSerGlnValTyrAlaLeuGluGlyCysTyrValLeuAlaPro 225
QY 673 TGGCGGACCGTTTCGCCGAGATGATCAAGGTAATTGGTGATACGCCCGCAAGAGATG 732
    |||||
Db 226 CysAlaLeuValSerLysGluMetLeuGluMetLeuAlaAspTrnProGlnLysArgGln 245
QY 733 TTCCTCAAGCGCGCGCGGTTTTCCTATGATTTTCGGGCGCGACGCGCGCGCTGGCC 792
    |||||
Db 246 LeuLeuLeuGluGlyGlyPheAlaGlnIlePheGlyProAspAlaLysProLeuCys 265
QY 793 GAGCGCTCCCGGAGACCGAAGAGGAGCTGCTGTCGCCGATATGACCTCGCATGATC 852
    |||||
Db 266 GluProPheProGluThrGlnGluGlyLeuLeuTyrAlaAspValaAspLeuGlyPheIle 285
QY 853 GCGTTGGCCAAGCGCGCGCGCATCCGCGCGCGCCACTATTCAAGCGCGCGCAAGTAAAGCGG 912
```

```

Db      286 GlyValAlaIysAlaIaIaTyAspProThrGlyHisTyrSerArgProAspValValArg    305
Qy      913 CTGCTGCTGGATCGACGTCCGGCCCAAGCGCTGCATCAGCTTGATGCC          960
Db      306 LeuLeuTrpAsnSerLysProAlaThrArgValHisSerPheGlnAla          321

RESULT 9
Q6RWS5_9ZZZZ
ID Q6RWS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
AC Q6RWS5;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE NitriIase (EC 3.5.5.7).
GN ORFNames=BD7611;
OS uncultured organism.
OC unclassified; environmental samples.
ON NCBI_TaxId=155900;
RX [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitriIase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487427; AAR97374.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitriIase activity; IEA.
DR GO; GO:0016810; F:hydroIase activity, acting on carbon-nitrog. . ; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydratase.
DR InterPro; IPR003010; Ntlse/CNhdtse.
DR Pfam; PF00795; CN_hydroIase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
DR HydroIase.
SQ SEQUENCE 345 AA; 37242 MW; 5F8A3CFE530D2F1D CRC64;

Alignment Scores:
Pred. No.: 3.09e-59 Length: 345
Score: 1125.00 Matches: 212
Percent Similarity: 77.1% Conservative: 44
Best Local Similarity: 63.9% Mismatches: 72
Query Match: 59.8% Indels: 4
DB: 2 Gaps: 1

US-09-751-299-1 (1-1041) x Q6RWS5_9ZZZZ (1-345)

Qy      19 AAGTATCGCGCGCGCGGTGACAGCCGCGCGTGTTCCTCGATCTGCACCGACAGTC       78
Db      7 ArgTyrArgValAlaIaIaValGlnAlaIaProGluPheLeuAsnLeuGlnAlaThrVal     26
Qy      79 GAGAAAGCATCGCGCTGATCGAGCAGGCGGCCAAGCAGACGATGCCCTGATTCGATTTC       138
Db      27 AsplysThrIleAlaIeuleIleGluIlaIaIaArgGlyGlyAlaSerIleuIleAlaPhe       46
Qy      139 CAGAGACTTGATTCGCGGTATCCCTTTTGATATGGCTGGCGCGCGCTTGGGGC         198
Db      47 ProGluThrTripleProGlyTyrProTriPheAlaTriPleuGlyAlaProIleTriPgly       66
Qy      199 ATGCGCTTCGTCACAGCGTATTTCGAATAATTCGCTCGTGGCGGACGACGACATGGCACG       258
Db      67 MetLysPheIleGlnAlaTyrHisAspAsnSerMetValIleaspGlyAlaGlnPheGlu       86
Qy      259 GCCCTGGCGGATGCGGCGCGCGCGCCACGCGCATGCTGCTGGCGCGGCTATAGCGAGCGC       318
Db      87 ArgIleAlaGlnAlaIaIaSerArgCysAsnIleThrValValLeuGlyPheSerGluLys       106
Qy      319 GCGGGCGGACGCTCTATATGGGCCAGGGCGATCTTGGCCCCGATGGCGATCGATCGCC        378

```

Db	10	AspAlaGlySerLeuTyrIleAlaGlnAlaIleLeuSerProGluGlyThrIleAla	126
QY	379	GGCGCCGCGCAAGCTCAAGCCTTACCATTGCGGAGCGCACCGTGTTCGGCGAGGAGACGGC	438
Db	127	ThrArgArgLysLeuLysProThrHisValGluArgAlaIlePheGlyGluGlyAspGly	146
QY	439	AGCCATTCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGCGCGCTCTGTGCTGGAG	498
Db	147	SerAspLeuAlaValHisAspThrLysLeuGlyArgValGlyAlaLeuCysTrpGlu	166
QY	499	CACATCCAGCCATTGTGGAATACGCCATGTACGCCGCCGACCAACAGGTTCACCTCGCG	558
Db	167	HisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisIleAla	186
QY	559	TCGTGGCCGAGCTTCAGCCTCTATCGCGGCATGCGCCTATGCGCTCGGACCGAGTCAAT	618
Db	187	AlaTrpProSerPheSerLeuTyrValAspAlaAlaTyrAlaLeuGlyProGluValAsn	206
QY	619	ACCGCCGCAAGCCAGATCTACCGCGGTGAGGCGCGCTGTACGTGCTGCGCTGCGCG	678
Db	207	AsnAlaAlaSerArgLeuTyrAlaValGluGlyGlnCysPheValValAlaProCysAla	226
QY	679	ACCGTTTCGCGGAGATGATCAAGGTATTGTGTGATACGCCCGACAGAGAGATGTTCTC	738
Db	227	ThrValSerGlnLysMetIleAspMetLeuCysGluThrProGluGlnAlaLeuLeu	246
QY	739	AAGCGCCGCGCGCGGTTTGGCATGATTTTCGGGCGCCGACGCGCGCCCTGGCCGACCG	798
Db	247	LysProGluGlyGlyHisAlaGlnIleTyrGlyProAspGlyArgSerLeuAlaAspPro	266
QY	799	CTCCCGGAGACCGAAGAGGAGCTGTGTGTCGCCGATATCGACCTCGCATGATCGCGTTG	858
Db	267	LeuProProAspAlaGluGlyLeuLeuTyrAlaAspIleAspLeuAlaIleThrLeu	286
QY	859	GCCAAGCGCGCGCGCATCCGGCGGCGCCACTATTACGCGCCGACGTAACGCGGCTGCTG	918
Db	287	AlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgProAspValThrGlnLeuLeu	306
QY	919	CTGGATCGACGTCCGCGCCCAACGCGCTGTCACGCTTGATGCCGCAATTCGAACCGCAAAAC	978
Db	307	LeuAspArgAsnProLysProArgValValHisAla-----LysProGluGln	322
QY	979	GAGGACAAAGGCGCAGCGCGCGCGCTGCGGTG 1014	
Db	323	SerAlaAsnAsnSerSerProGlyMetArgAlaVal 334	
RESULT 10			
Q6RWE3_9ZZZZ PRELIMINARY; PRT; 334 AA.			
ID	Q6RWE3_9ZZZZ		
AC	Q6RWE3;		
DT	05-JUL-2004 (TREMBLrel. 27, Created)		
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)		
DDE	Nitrilase (EC 3.5.5.7).		
GN	ORFNames=BD7051;		
OS	uncultured organism.		
OC	unclassified; environmental samples.		
OX	NCBI_TaxID=155900;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RX	PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;		
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,		
RA	Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,		
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,		
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.;		
RT	"Exploring nitrilase sequence space for enantioselective catalysis.";		
RL	Appl. Environ. Microbiol. 70:2429-2436 (2004).		
DR	EMBL; AY487559; AAR97506.1; -; Genomic DNA.		
DR	GO; GO:0018762; F:aliphatic nitrilase activity; IEA.		
DR	GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.		
DR	GO; GO:0006807; F:nitrogen compound metabolism; IEA.		
DR	InterPro; IPR000132; N/CN hydratase.		
DR	InterPro; IPR003010; NtIse/CNhydrtase.		





```
Db      87 AlaGluLysIleArgLysAlaValLysArgAlaLysLeuThrAlaValIleGlyValSer 106
      313 GAGCGCGCGGCGGAGCGCTCTATATGGCCAGGCGATCTTGGCCCGGATGGCATCTG 372
      107 GluArgAspGlyGlySerLeuTyrIleGlyIleThrLeuIleGlyProAspGlyGluThr 126
QY      373 ATCGCCGCGCGCGCAAGCTCAAGCTTACCCATGGCGAGCGCACCGGTGTCGGCAGGGA 432
Db      127 IleAlaLysArgArgLysLeuArgProThrHisAlaGluArgThrValPheGlyGluGly 146
QY      433 GACGGCAGCCATCTCGCGGTGCAAGATACCGCC---ATCGGGCGCTCGCGCGCTCTGT 489
Db      147 AspGlySerAspLeuAlaValHisAspArgAlaAspValGlyArgLeuGlyAlaMetCys 166
QY      490 TGCTGGAGACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCGCAACAAGCTC 549
Db      167 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnVal 186
QY      550 CACGTGCGCTCGTGGCGGAGCTTCAAGCTCTATCCGGCGCATGGCCCTATGCGCTCGAACC 609
Db      187 HisValGlyAlaTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyTrp 206
QY      610 GAGGTCAATACCGCGCGCAAGCCAGATCTACCGGTCGAGGGCGGCTGCTACGTGCTGGCG 669
Db      207 GluValAsnAsnAlaAlaSerLysValTyrAlaValGluGlySerCysPhePheLeuGly 226
QY      670 TCGTGGCGGACCGCTTTCGCGGAGATGATCAAGGTATTGTGATACGCCCGCAACAAGAG 729
Db      227 ProGlyAlaValValSerGlnAlaMetIleAspGluLeuCysAspSerProGluLysHis 246
QY      730 ATGTTCTCAAGCGCGCGCGGGGTTTGGCATGATTTTCGGGCGCGCGCGCGCGCTG 789
Db      247 AlaPheLeuHisAlaGlyGlyGlyHisAlaValIleTyrGlyProAspGlySerSerLeu 266
QY      790 GCCGAGCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCGCGCATATGACCTCGGCATG 849
Db      267 AlaAspLysLeuProProAspGlnGlyIleLeuTyrAlaAspIleAspLeuGlyMet 286
QY      850 ATCGCGTTGGCCCAAGCGCGCGGATCCGGCGGGCCACTATTACGGCGCGCAAGTAAAC 909
Db      287 IleGlyValAlaLysAsnAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 306
QY      910 CGGCTGCTGCTGATCGACGTCGCGGCCCAAGCGCTC 945
Db      307 ArgLeuLeuAsnThrSerArgAlaAsnArgVal 318

RESULT 12
Q6RWG6_9ZZZZ PRELIMINARY; PRT; 338 AA.
ID Q6RWG6_9ZZZZ
AC Q6RWG6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NitriLase (EC 3.5.5.7).
GN ORFNames=BD5279;
OS unclutered organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitriLase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487536; AAR97483.1; -; Genomic_DNA.
DR GO; GO:0018762; F:aliphatic nitriLase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR00132; N/CN_hydrolase.
```

```
DR InterPro; IPR003010; Ntlase/CNhydrlse.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW Hydrolase.
SQ SEQUENCE 338 AA; 36483 MW; B9EBD7FC2164988D CRC64;

Alignment Scores:
Pred. No.: 3.77e-56 Length: 338
Score: 1073.50 Matches: 213
Percent Similarity: 77.9% Conservative: 30
Best Local Similarity: 68.3% Mismatches: 66
Query Match: 57.1% Indels: 3
DB: 2 Gaps: 2

US-09-751-299-1 (1-1041) x Q6RWG6_9ZZZZ (1-338)
QY 19 AAGTATCGCGCGCGCGCGGTGCAAGCGCGCGCGGTGTTCTTCGATCTGACCGCACAGTC 78
Db 7 LysTyrArgValAlaAlaValGlnAlaAlaProAlaTyrLeuAspLeuAspArgSerIle 26
QY 79 GAGAAAGCGATCGGCGCTGATTCAGACAGCGCGCCCAAGCAGCAGCTGCGCTGATTCGATT 138
Db 27 AspLysAlaIleAlaLeuIleGluGluAlaAlaAlaAsnGlyAlaArgLeuIleAlaPhe 46
QY 139 CCAAGACTTGATTCGCCGCTATCCCTTTGGATATGCTGGCGCGCGCGCTTG--- 195
Db 47 ProGluValPheIleProGlyTyrProTyrPheHisIleTrpLeuAspSerProAlaTrpAla 66
QY 196 ---GGCATGCGCTTCGTCAGCGCTATTTCGAGAATTCGCTGCGCGCGCACAGCAG 252
Db 67 IleGlyArgGlyPheValGlnArgTyrPheAspAsnSerLeuAlaTyrAspSerProGln 86
QY 253 TGGCAGGCCCTGGCGGATGCGGCCCGCGCCAGCAGCATGTCGTGCGCGCTATAGC 312
Db 87 AlaGluArgLeuArgAlaAlaValArgLysAlaArgLeuThrAlaValIleGlyLeuSer 106
QY 313 GAGCGCGCGCGCGCGGAGCTCTATATGGCCAGGCGCATCTTGGCGCGCGCGCGCATCTG 372
Db 107 GluArgSerGlyGlySerLeuTyrIleAlaGlnTrpLeuValGlyProAspGlyGluThr 126
QY 373 ATCGCCGCGCGCGCAAGCTCAAGCTTACCCATGGCGGAGCGCACCGGTGTCGGCAGGGA 432
Db 127 IleAlaLysArgArgLysLeuArgProThrHisAlaGluArgThrValTyrGlyGluGly 146
QY 433 GACGGCAGCCATCTCGCGGTGCAAGATACCGCC---ATCGGGCGCTCGCGCGCTCTGT 489
Db 147 AspGlySerAspLeuAlaValHisAspArgProAspIleGlyArgLeuGlyAlaLeuCys 166
QY 490 TGCTGGAGACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCGCAACAAGCTC 549
Db 167 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnVal 186
QY 550 CACGTGCGCTCGTGGCGGAGCTTCAAGCTCTATCGCGGCGCATGGCCCTATGCGCTCGAACC 609
Db 187 HisValAlaSerTrpProSerPheSerLeuTyrAspProPheAlaProAlaLeuGlyAla 206
QY 610 GAGGTCAATACCGCGCGCAAGCCAGATCTACCGGTCGAGGGCGGCTGCTACGTGTCGGCG 669
Db 207 GluValAsnAsnAlaAlaSerArgValTyrAlaValGluGlySerCysPheValLeuAla 226
QY 670 TCGTGGCGGACCGCTTTCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATG 729
Db 227 ProGlyAlaThrValSerGlnAlaMetIleAspGluLeuCysAspArgProAspLysHis 246
QY 730 ATGTTCTCAAGCGCGCGCGGGTTCGCAATGATTTTCGGGCGCGCGCGCGCGCTG 789
Db 247 AlaLeuLeuHisAlaGlyGlyGlyPheAlaAlaIleTyrGlyProAspGlySerSerLeu 266
QY 790 GCCGAGCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCGCGCATATGACCTCGGCATG 849
Db 267 AlaGluLysLeuAlaProAspGlnGlyLeuLeuTyrAlaAspIleAspLeuGlyAla 286
```



[illegible]

```

Db      101 ValSerLeuGlyPheAlaGluArgAspGlyThrLeuTyrIleAlaGlnTrpLeuIle 120
QY      355 GGGCCCGATGGCCGATCTGATCGCCCGCCGCCAAGCTCAAGCCTTACCATTGCGGAGCGC 414
Db      121 GlyProAspGlyGlnThrIleSerArgArgGlySLeuLysProThrHisValGluArg 140
QY      415 ACCGTGTTCCGGAGGGAGAGACGGCAGCCATCTCGCGGTGCACGATACGCCCATCGGGCGC 474
Db      141 ThrValPheGlyGluGlnLysAspGlySerAspLeuSerValHisAspThrAlaLeuGlyArg 160
QY      475 CTCGGCGCGCCTCTGTGCTGGAGACACATCCAGCCATTGTGGAATAAGCCATGTACGCC 534
Db      161 ILeGlySerLeuCysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrAla 180
QY      535 GCCGACGAACAGGTCCACGTCGCGCTCGTGGCCGAGCTTCAGCCTTATTCGGCGCATGGCC 594
Db      181 GlnAsnGluGlnIleHisIleGlyAlaTrpProSerPheSerLeuTyrGlnProPheAla 200
QY      595 TATGCGCTCGGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGC 654
Db      201 AsnAlaLeuSerProGluValAsnIleAlaValSerArgValTyrAlaValGluGln 220
QY      655 TGCTACGTGTCGCGCTGCTGCGCGACCGCTTTCGCGGAGATGATCAAGTATTGTGTGAT 714
Db      221 CysPhePheLeuAlaProCysAlaThrValSerAspAlaMetIleGluThrLeuCysAsp 240
QY      715 ACGCCCGACAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGGCC 774
Db      241 ThrProGluLysGlnGlyLeuIleArgAlaGlyGlyHisAlaIlePheGlyPro 260
QY      775 GACGCGCGCGCCTGGCCGAGCGCTCCCGAGACCGAAGAGGAGCTGCTGTCGGCAT 834
Db      261 AspGlySerLeuLeuThrProThrValAlaAspThrTyrGluGlyLeuLeuTyrAlaGlu 280
QY      835 ATCGACCTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGCGCGGCCACTATTCA 894
Db      281 LeuAspLeuGlyValIleSerIleAlaLysSerAlaAlaAspProAlaGlyHisTyrSer 300
QY      895 CGGCGCGACGTAAACGGCGGTGCTGCTGCATCGACGTCCGGCCCAACCGCTGCTCAGCCTT 954
Db      301 ArgProAspValThrArgLeuLeuLeuAsnGlnThrProSerLysArgValGlnAsnMet 320
QY      955 GATGCCGATTCGAACCGCAAAACGAGACAAAGGC 990
Db      321 ValLeuProLeuGluThrValThrGluProGluGly 332

RESULT 14
Q6RMF7_9ZZZZ PRELIMINARY; PRT; 335 AA.
AC Q6RMF7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NitriLase (EC 3.5.5.7).
GN ORFNames=BD5326;
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitriLase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
EMBL: AY487545; AAR97492.1; -; Genomic_DNA.
DR GO; GO:0018762; F:aliphatic nitriLase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN_hydrolase.

```

DR InterPro; IPR003010; Ntlase/cnhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 335 AA; 36902 MW; COABIDF5F963AC6A CRC64;  
  
Alignment Scores:  
Pred. No.: 7.52e-56 Length: 335  
Score: 1068.50 Matches: 209  
Percent Similarity: 74.9% Conservative: 42  
Best Local Similarity: 62.4% Mismatches: 79  
Query Match: 56.8% Indels: 5  
DB: 2 Gaps: 2  
  
US-09-751-299-1 (1-1041) x Q6RMF7\_92ZZZ (1-335)  
  
QY 1 ATGTGAGAGCCCATGACGAATGATCGCGCGCGGTGACAGCCGCGGTTCCTC 60  
DB 1 MetSerIleThrHisProLysPheLysAlaIaValGlnAlaIaIaProValPheLeu 20  
QY 61 GATCTCGACCGACAGTCGAGAAAGCGATCGCGCTGATCGAGCAGCGCGCAAGCAGAC 120  
DB 21 AspLeuAspGlySerValAsnLysAlaIaLeuLeuIleAspGluAlaIaIaIaGly 40  
QY 121 GTGCGCTGATCGCATTCACAGAGACTTGATCCCGCTATCCCTTTGATATGCTG 180  
DB 41 AlalysLeuIleAlaPheProGluThrPheIleProGlyTyrProIleTyrIleTyrLeu 60  
QY 181 GCGCGCGCGCTTG-----GGCATGCGCTTGTCCAGCGCTATTTGAGAATTCGCTC 234  
DB 61 GlySerProAlaTyrAlaLeuGlyArgGlyPheValGlnArgTyrPheAspAsnSerLeu 80  
QY 235 GTGCGCGCGACAGCAGTGGCAGCGCCCTGGCGGATGCGCGCGCGCGCAAGCATGCT 294  
DB 81 GlnTyrAspSerProGlnAlaAspArgLeuArgLysAlaIaArgArgAsnSerIleThr 100  
QY 295 GTGCGCGCGCTATAGCAGCGCGCGCGCGCGCGCTTATATGCGCGCAAGCATCTTC 354  
DB 101 ValValLeuGlyLeuSerGluArgAspGlyGlySerLeuTyrIleAlaGlnTyrLeuIle 120  
QY 355 GCGCGCGATGCGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCTCAAGCATCGCGCGC 414  
DB 121 GlyProAspGlyGluThrIleAlaGlnArgArgLysLeuArgProThrHisGlyGluArg 140  
QY 415 ACCGTGTCGCGCGGAGAGAGCGCAGCCATCTCGCGTGACGATACCGCATCGCGCGC 474  
DB 141 ThrValPheGlyGluGlyAspGlySerAspLeuValValHisGlnThrGluLeuGlyArg 160  
QY 475 CTGCGCGCGCTCTGTGTCGAGCAGCATCCAGCATCTGCGAATAACGCCATGTACGCC 534  
DB 161 LeuGlyAlaLeuAsnCysTyrGluAsnIleLeuSerLeuAsnLysTyrValMetTyrSer 180  
QY 535 GCGGACGACAGGTCCACGTCGCGCTGCGCGCGCGAGCTTCAGCCTCTATCGCGCATGCC 594  
DB 181 GlnHisGluGlnValHisValAlaSerTyrProSerPheSerThrTyrGluProPheAla 200  
QY 595 TATGCGCTCGGACCGGAGGTCAATACCGCCGCAAGCCAGCATCTACCGCGTCGAGCGCGC 654  
DB 201 HisAlaLeuGlyTyrGluValAsnAsnAlaIleSerGlnValTyrAlaValGluGlyGly 220  
QY 655 TGCTACGTGCTGGCGCTGTCGCGCGACCGCTTTCGCGGAGATGATCAAGGTATTGCTGAT 714  
DB 221 CysPheValLeuAlaProCysSerThrIleSerGluGluMetIleAlaGluLeuCysAsp 240  
QY 715 ACGCCGACACAGAGATGTTCCCTCAAGCGCGCGCGCGGTTCGATGATTTCCGGCGCC 774  
DB 241 ThrProAspLysPheGluLeuThrHisAlaGlyGlyGlyHisAlaIleIleTyrGlyPro 260  
QY 775 GACGCGCGCGCGCTGCGCGAGCGCGCTCCCGGAGACCGCAAGAGGAGCTGCTGTCGCGCAT 834  
DB 261 AspGlyArgAlaLeuCysGlyLysLeuProGluAsnGlnGluGlyLeuLeuTyrAlaGlu 280

QY 835 ATGACCTCGCATGATCGCGTGGCCCAAGCGCGCGCGCATCCGCGCGGCACTATTCA 894  
DB 281 IleAspLeuGlyValIleSerMetAlaLysSerAlaMetAspProValGlyHisTyrSer 300  
QY 895 CGGCGCGAGTAAACGCGCGTGTCTGATGACGCTCCGCGCGCAAGCGCTGTCACGCTT 954  
DB 301 ArgProAspValTyrArgValLeuPheAsnLysIleProAlaLysArgIleGluHisPhe 320  
QY 955 GATGCGCGATTCGAACCGCAAAACAGAGACAAAGGCGGACGCGCC 999  
DB 321 AsnLeu-----ProLeuAspGluGlnAlaGlyGluGluPro 332  
  
RESULT 15  
ID Q5YUM5\_NOCFA PRELIMINARY; PRT; 336 AA.  
AC Q5YUM5;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Putative nitrilase.  
GN OrderedLocusNames=nfa32690;  
OS Nocardia farcinica.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Nocardiaceae; Nocardia.  
OX NCBI\_TaxID=37329;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=IFM 10152;  
RX PubMed=1546710; DOI=10.1073/pnas.0406410101;  
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,  
RA Shiba T., Hattori M.;  
RT "The complete genomic sequence of Nocardia farcinica IFM 10152.";  
RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).  
DR EMBL; AP006618; BAD58116.1; -, Genomic\_DNA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR DR; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntlase/cnhydase.  
DR InterPro; IPR002332; PII GlnB UMP\_S.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 336 AA; 36091 MW; D24921911BD0C2BB CRC64;  
  
Alignment Scores:  
Pred. No.: 7.52e-56 Length: 336  
Score: 1068.50 Matches: 202  
Percent Similarity: 75.2% Conservative: 47  
Best Local Similarity: 61.0% Mismatches: 79  
Query Match: 56.8% Indels: 3  
DB: 2 Gaps: 2  
  
US-09-751-299-1 (1-1041) x Q5YUM5\_NOCFA (1-336)  
  
QY 22 TATGCGCGCGCGGTGACGAGCGCGCGGTTCCTCGATCTCGACCGCAGTCGAG 81  
DB 7 PheArgAlaAlaAlaValGlnAlaIaIaProValTyrPheAspGlyAlaIaIaThrValAsp 26  
QY 82 AAAGCATCGCGCTGATCGAGCAGCGCGCAAGCAGAGCAGTGCCTGATGCAATCCCA 141  
DB 27 LysCysValAlaLeuIleGluGluAlaIaIaAspAsnGlyAlaIaIaLeuIleAlaPhePro 46  
QY 142 GAGACTTGATTCGCGCTATCCCTTTTGATATGCTGGCGCGCGCTTGGGGCATG 201  
DB 47 GluThrPheValProGlyTyrProTyrTyrPheLeuTyrPheAspSerProAlaTyrGlyMet 66  
QY 202 CGCTTCGTCAGCGCTATTTGAGAATTGCTGTCGCGCGCAGCAGCAGTGGCAGGCC 261  
DB 67 GlnPheValAlaArgTyrPheAspAsnSerLeuAlaLeuAspGlyProLeuPheAlaArg 86

```
QY      262 CTGGCGGATGCGCGCCGCCACGCGCATGTCGTGGCCGCTATAGCAGCGCGC 321
      |||  ::||| ||| ||| |||  ::| ||| ||| ||| ||| |||
Db      87  LeuArgGluAlaAlaArgSerAlaIleThrValValThrGlyHisSerGluArgAsp 106
      |||  ::||| ||| ||| ||| |||  ::| ||| ||| ||| ||| |||
QY      322 GCGCGCAGCCTCTATATGCGCCAGCGCATCTTCGCGCCCGATGGCGATCTGATCGCCCGC 381
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      107 GlyGlySerLeuTyrMetGlyGlnAlaIleIleGlyAlaAspGlyGluValLeuAlaAla 126
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      382 CGCCGCAAGCTCAAGCCTTACCCTATCGGAGCGCACCGCTGTTCGGCGAGGAGACGGCAGC 441
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      127 ArgArgLysLeuLysProThrHisValGluArgThrValPheGlyGluSerAspGlySer 146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      442 CATCTCGCGGTGCACGATACCGCCATCGGCGCGCTCGCGCTCTGTGCTGGAGAC 501
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      147 AsnLeuThrValValAspThrGluLeuGlyArgLeuGlyAlaLeuCysCysTrpGluHis 166
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      502 ATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGACGACAGTCCACGTCCGCTCG 561
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      167 LeuGlnProLeuThrLysTyrAlaMetTyrSerGlnHisGluGlnIleHisValAlaAla 186
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      562 TGGCCGAGCTTCAGCCTCTATCGCGCATGGCCTATGCGCTCGGACCGGAGTCAATACC 621
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      187 TrpProSerPheSerValTyrArgGlyAlaAlaIleTyrAlaLeuGlyProGluValAsnThr 206
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      622 GCCGCAAGCCAGATCTACGCGGTGAGGGCGCGGTGCTACGTGCTGGCGTGCAGCACC 681
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      207 GlyAlaAlaArgGlnTyrAlaValAlaGluGlyGlnCysPheValLeuSerProCysAlaVal 226
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      682 GTTTCGCCGAGATGATCAAGTATTGTGATACGCCCGACAGAGAGATGTCCTCAAG 741
      ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      227 IleAspGluAlaGlyValGluLeuPheCysAspThrProAlaLysArgGluLeuLeu 246
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      742 GCCGCGCGCGGTTTGCATGATTTTCGGGCCCGACGGCCCGCCCTGGCCGACCGCTC 801
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      247 ProGlyGlyGlyPheAlaGlnIleTyrGlyProAspGlyArgGluLeuGlyThrAlaLeu 266
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      802 CCGAGACCCGAAGAGGACTGCTGTGTCGCCGATATCGACCTCGGCATGATCGCGTTGGCC 861
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      267 ProGluThrGluGluGlyLeuValTyrAlaAspLeuGluAlaSerAlaValAlaValAla 286
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      862 AAGCGCGCGCGCATCCGGCGGGCCACTATTCACGGCCGACGTAACGGCGCTGCTGTG 921
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      287 LysSerAlaAlaAspProValGlyHisTyrSerArgProAspValLeuGlnLeuLeuTrp 306
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      922 GATCGACGTCCGCCCAA-----CGCGTCGTCACGCTTGATGCCGATTCGAACCGCAA 975
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      307 AspProArgProArgSerValValArgGlnValAlaLeuSerValAlaSerProAlaGlu 326
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      976 AACGAGACAAGGGCGACGCGCCCGCGCTGCGC 1008
      ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      327 SerAlaAspAspAlaGlu---ProAlaValArg 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: April 27, 2006, 01:35:33  
Job time : 263.071 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:06:26 ; Search time 34.8013 Seconds  
(without alignments)  
2628.598 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1881  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 segs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/absas/ABSSWEB\_spool/US09751299/runat\_26042006\_090032\_18619/app\_query.fasta\_1  
-DB=A\_Geneseq -QFWT=faстан -SUFFIX=rag -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=absas02p  
-USER=US09751299\_@CGN\_1\_1\_476\_@runat\_26042006\_090032\_18619 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_21:\*

1:	geneseqp1980s:*
2:	geneseqp1990s:*
3:	geneseqp2000s:*
4:	geneseqp2001s:*
5:	geneseqp2002s:*
6:	geneseqp2003as:*
7:	geneseqp2003bs:*
8:	geneseqp2004s:*
9:	geneseqp2005s:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1806	96.0	346	4	AAE05488	Aae05488 Nitrilase
2	1806	96.0	346	7	ADC24117	Adc24117 Protein s
3	1806	96.0	346	8	ADH36218	Adh36218 Chemical
4	1806	96.0	346	8	ADG93917	Adg93917 Nitrilase
5	1806	96.0	346	8	ADi62516	Adi62516 Nitrilase
6	1806	96.0	346	8	ADi64637	Adi64637 Nitrilase
7	1464	77.8	341	7	ADC23767	Adc23767 Protein s
8	1464	77.8	341	8	ADH35868	Adh35868 Chemical
9	1464	77.8	341	8	ADG93569	Adg93569 Nitrilase

10	1464	77.8	341	8	ADi62166	Adi62166 Nitrilase
11	1464	77.8	341	8	ADi64287	Adi64287 Nitrilase
12	1209	64.3	348	7	ADC23749	Adc23749 Protein s
13	1209	64.3	348	8	ADH35850	Adh35850 Chemical
14	1209	64.3	348	8	ADG93551	Adg93551 Nitrilase
15	1209	64.3	348	8	ADi62148	Adi62148 Nitrilase
16	1209	64.3	348	8	ADi64269	Adi64269 Nitrilase
17	1165	61.9	353	7	ADC23945	Adc23945 Protein s
18	1165	61.9	353	8	ADH36046	Adh36046 Chemical
19	1165	61.9	353	8	ADG93747	Adg93747 Nitrilase
20	1165	61.9	353	8	ADi62344	Adi62344 Nitrilase
21	1165	61.9	353	8	ADi64465	Adi64465 Nitrilase
22	1155	61.4	353	7	ADC23965	Adc23965 Protein s
23	1155	61.4	353	8	ADH36066	Adh36066 Chemical
24	1155	61.4	353	8	ADG93767	Adg93767 Nitrilase
25	1155	61.4	353	8	ADi62364	Adi62364 Nitrilase
26	1155	61.4	353	8	ADi64485	Adi64485 Nitrilase
27	1148	61.0	333	7	ADC24083	Adc24083 Protein s
28	1148	61.0	333	8	ADH36184	Adh36184 Chemical
29	1148	61.0	333	8	ADG93883	Adg93883 Nitrilase
30	1148	61.0	333	8	ADi62482	Adi62482 Nitrilase
31	1148	61.0	333	8	ADi64603	Adi64603 Nitrilase
32	1132.5	60.2	358	7	ADC23903	Adc23903 Protein s
33	1132.5	60.2	358	8	ADH36004	Adh36004 Chemical
34	1132.5	60.2	358	8	ADG93705	Adg93705 Nitrilase
35	1132.5	60.2	358	8	ADi62302	Adi62302 Nitrilase
36	1132.5	60.2	358	8	ADi64423	Adi64423 Nitrilase
37	1130	60.1	354	7	ADC23835	Adc23835 Protein s
38	1130	60.1	354	8	ADH35936	Adh35936 Chemical
39	1130	60.1	354	8	ADG93637	Adg93637 Nitrilase
40	1130	60.1	354	8	ADi62234	Adi62234 Nitrilase
41	1130	60.1	354	8	ADi64355	Adi64355 Nitrilase
42	1125	59.8	345	7	ADC24067	Adc24067 Protein s
43	1125	59.8	345	8	ADH36168	Adh36168 Chemical
44	1125	59.8	345	8	ADG93868	Adg93868 Nitrilase
45	1125	59.8	345	8	ADi62466	Adi62466 Nitrilase

ALIGNMENTS

RESULT 1	
AAE05488	
ID	AAE05488 standard; protein; 346 AA.
XX	
AC	AAE05488;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Nitrilase protein #1.
XX	
KW	Nitrilase; enantiomer; alpha-substituted carboxylic acid; stereoselective; hydrolysis; amino nitrile; cyanohydrin.
XX	
OS	Unidentified.
XX	
PN	WO200148175-A2.
XX	
PD	05-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US035555.
XX	
PR	29-DEC-1999; 99US-0173609P.
XX	
PA	(MADD/) MADDEN D.
XX	
PI	Madden M, Weiner DP, Chaplin JA;
XX	
DR	WPI, 2001-465211/50.
XX	
PT	N-PSDB; AAD11166.
	Producing alpha-substituted carboxylic acid enantiomers by contacting

PT aldehyde or ketone with cyanide, ammonia compound or its salt or amine  
PT and hydrolyzing the resulting amino nitrile or cyanohydrin with  
PT Nitrilase.  
XX  
PS Claim 29; Page 84-85; 87pp; English.  
XX  
CC The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolyzing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a nitrilase protein #1  
XX  
SQ Sequence 346 AA;

Alignment Scores:  
Pred. No.: 1.53e-148 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x AAE05488 (1-346)

QY 1 ATGTCGAGAGCCCATGACGAAGTATCGCGGCGCGGTGAGAGCCGCGCGGTTCCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATCTGACCGGACAGTCGAGAAAGCGATCGGCTGATCGAGAGCGGCGCAAGAGGAC 120  
Db 21 AspleuAspArgThrValGlnLysAlaIleGlyLeuIleGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCACAGACTTGATTCGCGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60  
QY 181 GCGCGCGCGGCTTGCGCATGCGCTTCGCCAGCTATTTCGAGAATTGCTGCTGCC 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCGCTGGCGATGCGGCGCGCCAGCGCATGCTGCTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCAGAGCGCGCGCGCGCAGCCTCTATATGGGCCAGCGCATCTTCGGCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGGATCTGATCGCGCGCGCGCGCAGCTCAAGCCTAACCCATGCGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGCGAGGAGAGACGCGCAGCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGAGACACATCCAGCCATTTGCGAAATACGCCATGTACCGCGCGAC 540  
Db 161 AlaLeuCySerTrpGlnHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTGCGCTGCGCGCAGACTTCAGCCTCTATCGCGGCGCATGCGCTTACG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetalArgAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGlnGlyGlyCysTyr 220  
QY 661 GTGCTGCGCTGCTGCGGACCGTTTCGCGCGAGATGATCAAGGTATTTGTTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240

QY 721 GACAAGAGATGTTCTCTCAAGCGCGCGCGGTTTGGCCATGATTTCCGGCCCGACGCG 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGGAGCCGCTCCCGGAGACCGAAGAGGAGTCTGGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATCGGCTTGCGCCAAAGCGCGCGCCGATCCGGCGGCGCACTATTTCAGGCCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGCGGCTGCTGTGATCGACGTCCGCGCCCAACGCGTCTGCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGGAACCGCAAAACGAGGACAAGGCGGACGCGCGCGCTGCGGTGCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 2

ADC24117  
ID ADC24117 standard; protein; 346 AA.  
XX  
AC ADC24117;

DT 18-DEC-2003 (first entry)  
XX

DE Protein sequence (SeqID 384) exhibiting nitrilase activity.

XX enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX enantiomer; chiral medicine.

XX Unidentified.

OS WO2003000840-A2.  
XX  
PN 03-JAN-2003.

XX PD 15-MAY-2002; 2002WO-US015983.  
XX  
PF 21-JUN-2001; 2001US-0300189P.

XX PR 30-JUL-2001; 2001US-0309006P.  
XX  
PR 22-JAN-2002; 2002US-0351336P.

XX PA (DIVE-) DIVERSA CORP.  
XX  
PA (MADD/) MADDEN D.

XX PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
XX PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.

XX DR N-PSDB; ADC24116.

XX PT Novel nitrilase polypeptide, useful for making (R) - or (S) -ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R) - or (S) -mandelic acid or (S) - or (R) -phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.

XX PS Claim 40; SEQ ID NO 384; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development

CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.

XX Sequence 346 AA;

Alignment Scores:

Pred. No.:	1.53e-148	Length:	346
Score:	1806.00	Matches:	346
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.0%	Indels:	0
DB:	7	Gaps:	0

US-09-751-299-1 (1-1041) x ADC24117 (1-346)

QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGCGAGCCCGCGGTTCCTC 60  
DB 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATCTCGACCGCAGCTCGAGAAAGCGATCGCCTGATCGAGCAGCGCCCAAGCAGAC 120  
DB 21 AspLeuAspArgThrValGluValAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATCCCGCTATCCCTTTGGATATGCTG 180  
DB 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCACGCGCTATTTCGAGAATTGCTCGTCCG 240  
DB 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGGCAGGCGCCCTGGCGGATGCGCGCGCGCCGCGCATGCATGCTG 300  
DB 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCAGCGCGCGCGCGCGCTTATATGCGCAGCGCATCTTCGCGCCC 360  
DB 101 AlaGlyTyrSerGluArgAlaGlyLysLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGCGCATCTGATCGCGCGCGCGCGCAGCTCAAGCCTTACCCAGTCCGAGCGCACCG 420  
DB 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGCGCAGGAGACGCGCAGCCATCTCCGCGGTGCACGATACCGCCATCGCGCGCTCG 480  
DB 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGGAGCACATCCAGCATGTGCAATATACGCCATGTACGCGCGCAC 540  
DB 161 AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCACGTCGCGTCTGTCGCGCAGCTTACGCTCTATCCGCGCATGGCCTATCG 600  
DB 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGAGCCGAGGTCAATACCGCGCAGCAGCATCTACGCGGTGCGAGCGCGCTGTAC 660  
DB 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCGCAGATGATCAAGGTATTGGTGAATACGCC 720  
DB 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGGAGATGTTCTCAAGGCCGCGCGCGGTTTGGCCATGATTTTGGGCCCGACGGC 780  
DB 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGAGCGCGCTCCCGAGAGCGAAGAGGAGTGTGCTGCGCGATATCGAC 840

DB 261 ArgAlaLeuAlaGluProLeuProGluThrGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGCGCATGATCGCGTGGCCAGACGCGCGCGCATCCGCGCGGCACTATTACGCGCC 900  
DB 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAAAGCGGCTGCTGTGATCGACGTCCGCGCCACGCGTGTGTCAGCTTGATGCC 960  
DB 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGACAGGGCGACCGCGCGCGCTGCGGTGCGCGAA 1020  
DB 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
DB 341 SerAlaAlaAlaGln 346

RESULT 3

ADH36218  
ID ADH36218 standard; protein; 346 AA.

XX ADH36218;

AC ADH36218;

DT 11-MAR-2004 (first entry)

XX Chemical process monitoring-related nitrilase protein sequence SeqID384.

KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.

XX Unidentified.

XX WO2003098187-A2.

XX 27-NOV-2003.

XX 15-MAY-2003; 2003WO-US015639.

XX 15-MAY-2002; 2002US-0380737P.

XX (DIVE-) DIVERSA CORP.

XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;

XX Mcquaid J, Stege J;

XX WPI; 2004-142708/14.

XX N-PSDB; ADH36217.

XX Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.

PS Claim 73; SEQ ID NO 384; 277pp; English.

XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.

XX Sequence 346 AA;

Alignment Scores:

Pred. No.:	1.53e-148	Length:	346
Score:	1806.00	Matches:	346
Percent Similarity:	100.0%	Conservative:	0

Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 8 Gaps: 0

US-09-751-299-1 (1-1041) x ADH36218 (1-346)

QY	1	ATGTCGAGACCCCATGACGAAGTATCGCGCGCGGTGCAGGCCGCGCGTGTCTC	60
DB	1	MetSerGluProMetThrLysTyrArgGlyAlaValGlnAlaIaIaProValPheLeu	20
QY	61	GATCTCGACCCGACAGTCAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAC	120
DB	21	AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGlnAlaIaLysGlnAsp	40
QY	121	GTGCGCTGATCGCATTCACAGAGACTTGATTCGCGCTATCCCTTTGATATGCGCTG	180
DB	41	ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu	60
QY	181	GGCGCGCGCTTGCGGCATGCGCTTCGTCCAGCGCTATTTCGAGAAATTCGCTCGTGC	240
DB	61	GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg	80
QY	241	GGCAGCAAGCAGTGGCAGGCGCGCGCGCGCTTATATATGCGCCAGCGCATGCTGTG	300
DB	81	GlySerLysGlnTrpGlnAlaLeuAlaAspAlaIaArgArgHisGlyMetHisValVal	100
QY	301	GCCGCTATAGCAGCGCGCGCGCGCGCGCTTATATATGCGCCAGCGCATCTTCGCGCC	360
DB	101	AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro	120
QY	361	GATGCGCATTCGATCGCGCGCGCGCGCGCAAGCTCAAGCTACCCATCGCGCAGCCGCTG	420
DB	121	AspGlyAspLeuIleAlaIaArgArgLysLeuLysProThrHisAlaGluArgThrVal	140
QY	421	TTGCGCGAGGAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGCGCGCTCGGC	480
DB	141	PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly	160
QY	481	GCGCTCTGTTGCTGGGAGACACATCCAGCCATTGTCGAATACGCCATGACGCGCGCAC	540
DB	161	AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaIaAsp	180
QY	541	GAACAGGTCCACGTCGCGTGTGCGCGAGCTTCAGCCTTATCGCGGACATGCGCTATGCG	600
DB	181	GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla	200
QY	601	CTCGGACCGGAGGTCAATACCGCGCGCAAGCATCTACGCGGTGAGGCGGCTGTAC	660
DB	201	LeuGlyProGluValAsnThrAlaIaSerGlnIleTyrAlaValGluGlyCysTyr	220
QY	661	GTGCTGCGCTGTGCGCGACCGCTTCGCGCGAGATGATCAAGTTATTGGTGATACGCC	720
DB	221	ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro	240
QY	721	GACAAGGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGCGCGACGCGC	780
DB	241	AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly	260
QY	781	CGCGCGCTGGCCGAGCGCGCTCCCGGAGACCGAAGAGGAGCTGTGTGCGCGATATGAC	840
DB	261	ArgAlaLeuAlaGluProLeuProGluThrGluGlyLeuLeuValAlaAspIleAsp	280
QY	841	CTCGGCATGATCGCGTTGGCCAGGCGCGCGCGATCCGCGCGGCCACTATTACGCGCC	900
DB	281	LeuGlyMetIleAlaLeuAlaLysAlaIaIaAspProAlaGlyHisTyrSerArgPro	300
QY	901	GACGTAACGCGGCTGCTGTGATCGACGTCCGCGCCAAAGCGTGTGATGATGCC	960
DB	301	AspValThrArgLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla	320
QY	961	GCAATTCGAACCGCAAAACGAGGACGAAGGCGCAGCGCGCGCTGCGCGTGTGCGGAA	1020
DB	321	AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu	340

QY	1021	AGCGCGCGCGCGCGCGCAG	1038
DB	341	SerAlaAlaAlaIaGln	346

RESULT 4

ADG93917  
ID ADG93917 standard; protein; 346 AA.

AC ADG93917;

DT 11-MAR-2004 (first entry)

DE Nitrilase enzyme amino acid sequence SegID384.

KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;

KW enantioselective transformation; enzyme.

OS Unidentified.

PN WO2003097810-A2.

PD 27-NOV-2003.

PF 15-MAY-2003; 2003WO-US015712.

PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

PA (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

DR WPI; 2004-090637/09.

DR N-PSDB; ADG93916.

PT New isolated or recombinant nucleic acid encoding a polypeptide having

PS Claim 44; SEQ ID NO 384; 295bp; English.

CC This invention is related to a novel isolated or recombinant nucleic acid

CC encoding a protein having nitrilase activity. Nitrilase's are capable of

CC converting nitrile's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated

CC nitrilase proteins of the invention have increased activity and stability

CC at increased pH and temperature when compared to those conventionally

CC used. In addition, the nucleic acid of the invention is useful for

CC screening enantioselective transformation. The present sequence is that

XX of a nitrilase enzyme of the invention.

SQ Sequence 346 AA;

US-09-751-299-1 (1-1041) x ADG93917 (1-346)

QY	1	ATGTCGAGACCCCATGACGAAGTATCGCGCGCGGTGCAGGCCGCGCGTGTCTC	60
DB	1	MetSerGluProMetThrLysTyrArgGlyAlaIaValGlnAlaIaProValPheLeu	20
QY	61	GATCTCGACCCGACAGTCAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAC	120
DB	21	AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGlnAlaIaLysGlnAsp	40
QY	121	GTGCGCTGATCGCATTCACAGAGACTTGATTCGCGCTATCCCTTTGATATGCGCTG	180



Db 41 ValArgLeuIleAlaPheProGluThrTripleProGlyTyrProPheTripletLeu 60  
QY 181 GGGCGCGCGCTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAATTCGCTCGCGC 240  
Db 61 GlYAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCTTGCGGATGGCGCCCGCCAGCGGATGCATGTCGTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCCGGCTATAGCAGCGCGCGCGCGAGCGCTCTATATGGCCAGGCGATCTTCGGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyLysSerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCTATGGCGAGCGACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCAGGAGACGGCAGCCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGGAGACACATCCAGCCATTTGCAATACGCCATGTACGCCCGCAG 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCACGTCGCGCTGCTGCGCGCGAGCTTCAGCCTCTATCGCGGATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGCGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValaGluGlyCysTyr 220  
QY 661 GTGCTGCGCTCGTGGCGCGACCGCTTTCGCGCGAGATGATCAAGGTATGGTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGGAGATGTTCTCAAGGCCGGCGCGGTTTGGCATGATTTCCGGGCCCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGAGCGCGCTCCCGGAGACCGAAGGAGGACTGCTGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTGCGCATGATCGCGTGTGGCCAAGCGCGCGCGATCCGGCGGCGCACTATTCACGGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGGCGTCTGCTGGATCGACGTCCGGCCCCAACGCGTGTCACGTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValaThrLeuAspAla 320  
QY 961 GCATTCGAACCGCAAAACGAGGACAAAGGGCGACCGCGCTGCGCGGTGGTGCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValaAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaGln 346  
RESULT 5  
ADI62516 ID ADI62516 standard; protein; 346 AA.  
XX AC ADI62516;  
XX DT 22-APR-2004 (first entry)  
XX DE Nitrlase polypeptide #192.  
XX KM Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;

KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipaeamic; enzyme.  
OS Unidentified.  
XX PN WO2003106415-A2.  
XX PD 24-DEC-2003.  
XX PF 13-JUN-2003; 2003WO-US018840.  
XX PR 13-JUN-2002; 2002US-0389317P.  
XX PR 28-JUN-2002; 2002US-0392944P.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX DR WPI; 2004-090821/09.  
XX DR N-PSDB; ADI62515.  
XX PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrlase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
XX acid and forming atorvastatin.  
XX PS Claim 41; SEQ ID NO 384; 253pp; English.  
XX CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrlase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrlase polypeptide obtained from an  
XX environmental sample.  
XX SQ Sequence 346 AA;  
SQ Alignment Scores:  
Pred. No.: 1.53e-148 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x ADI62516 (1-346)  
QY 1 ATGTGGAGCCCATGACGAGATTCGGCGCGCGGTGACGGCGCGCGGTTCCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCCAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGGCGCTGATGCAATTCACAGAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTripletProGlyTyrProPheTripletLeu 60  
QY 181 GGGCGCGCGCTTGGGCGATGCGCTTCGTCCAGCGCTATTTCGAGAATTGCTGTGCGC 240

Db 61 GLYAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGACGATGGCAGAGCCCTGGCGGATGCGGCCCGCCAGCCGCGCATGTCGTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCCGCTATAGCCAGCGCGCGCGCGGAGCCCTCTATATGGCCAGGCCGATCTTCGGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCCCGCGCGCGCAAGCTCAAGCCTACCCCATGGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGCAGCCCATCTCGCGGTGACGATACCGCCATCGGAGCGCCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTGTGGAGACACATCCAGCCATGTGCAATACGCGCATGTACGCGCGGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTGCGCTGTCGTCGCGAGCTTCAGCCTCTATCGCGGATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGAGCCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGTGCGCGGACCGCTTTCGCGCGAGATGATCAAGGTATTGTGATACGCCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGAGATGTTCTTCAAGGCCGCGCGGCTTTTGCCATGATTTTCGGGCCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATTCGCTTGCCCAAGGCGCGCGCGATCCGCGCGGCCACTATTACGCGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAAACGCGGCTGCTGTGATCGATCGATCCGCGCAACGCGTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGGACAAAGGCGACGCGCGCGCGCTGCGGTGTCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValAlaGlu 340  
QY 1021 AGCGCCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346  
RESULT 6  
ID ADI64637 standard; protein; 346 AA.  
XX AC ADI64637;  
XX DT 22-APR-2004 (first entry)  
XX DE NitriIase seq id 193.  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KW carboxylic acid; cyanohydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
, KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;

KW (R)-phenyl lactic acid derivative; % enantiomeric excess;  
KW % diastomeric excess; food additive; drug intermediate; nitrilase.  
XX Unidentified.  
OS US2004014195-A1.  
XX PN 22-JAN-2004.  
XX PD 15-MAY-2003; 2003US-00440523.  
XX PF 29-DEC-1999; 99US-0173609P.  
XX PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX WPI; 2004-121569/12.  
DR N-PSDB; ADI64636.  
XX PT Novel isolated or recombinant polypeptide having nitrilase activity,  
PT useful in production of food additives.  
XX PS Claim 46; SEQ ID NO 384; 105bp; English.  
XX CC The invention describes an isolated or recombinant polypeptide (I)  
CC comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its  
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining % enantiomeric excess or % diastomeric  
CC excess. (I) is useful in the production of food additives and drug  
CC intermediates. This is the amino acid sequence of a nitrilase of the  
CC invention.  
XX SQ Sequence 346 AA;  
SQ Alignment Scores:  
Pred. No.: 1.53e-148 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x ADI64637 (1-346)  
QY 1 ATGTCGAGACCCATGACGAAGTATCGCGCGCGGTGACAGGCCGCGCGGTGTCCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValAlaGlnAlaAlaProValPheLeu 20  
QY 61 GATCTGACCGCAGACTGCGAAGGCGATCGGCTGATCGAGCAGGCGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGATTCCAGAGACTTGATTCCGCGCTATCCCTTTGGATATGGCTG 180

Db 41 ValArgLeuIleAlaPheProGluThrTrrPleProGlyTyrProPheTrrPleTrrPleu 60  
QY 181 GGGCGCGCGGCTTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAAATTCGCTCGTGGCG 240  
Db 61 GlyAlaProAlaTrrPglyMetArgPheValGlnArgTyrPheGlyuAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGAGCCCTGGCGGATGGCGCCCGCCAGCGGATGCATGTCGTG 300  
Db 81 GlySerLysGlnTrrPglAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCCGGCTATAGCAGCGCGCGGGCGGAGCCTCTATATGGGCCAGGCGATCTTCGGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyLysSerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCAAGCTCAAGCCTACCAGCGGAGCGGACCGCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCAGGAGAGCGGAGCCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTGCTGGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540  
Db 161 AlaLeuCysCysTrrPgluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTTCACGTGCGGTGCGGAGCTTCAGCCTCTATCGCGGATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrrProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGAGCCGAGGTCAATACCGCGCGCAAGCGAGATCTACGCGGTGAGGGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGGTGGCGGACCGCTTCGCGCGGAGATGATCAAGGTATGGTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGGCCGCGCGGCTTTGGCCATGATTTCCGGGCCCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGGAGCCGCTCCCGGAGAGCGGAGGAGGACTGCTGCTGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATCGCTTGGCCAGGCGCGCGCGATCCGCGGGCCACTATTACGCGCCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGGCGGTGCTGCTGATCGAGTTCGCGCCCAACGCGTGTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTCGAACCGCAAAACGAGGACAGGAGCGGCGCGCGCTGCGCGTGCGGAGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaGln 346  
RESULT 7  
ADC23767  
ID ADC23767 standard; protein; 341 AA.  
XX AC ADC23767;  
XX 18-DEC-2003 (first entry)  
XX Protein sequence (SeqID 34) exhibiting nitrilase activity.  
DE enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX

KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO2003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR N-PSDB; ADC23766.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.  
PT  
XX  
PS Claim 40; SEQ ID NO 34; 560pp; English.  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydrins into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polypeptide is a protein sequence that exhibits nitrilase activity of the invention.  
CC  
XX  
SQ Sequence 341 AA;  
Alignment Scores:  
Pred. No.: 1.08e-118 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 7 Gaps: 0  
US-09-751-299-1 (1-1041) x ADC23767 (1-341)  
QY 1 ATGTGAGAGCCCATGACGAGTATCCGGCGCGGTGACAGCCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATTCGACCGCACAGTGCAGAAAGCATCGCGCTGATCGAGCGGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGAGAGACTTGATTCGCGGCTATCCCTTTGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrrPleProGlyTyrProLeuTrrPleTrrPleu 60  
QY 181 GGGCGCGCGGCTTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAAATTCGCTCGTGGCG 240  
Db 61 GlySerProAlaTrrPglyMetArgPheValGlnArgTyrPheGlyuAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATGGCGCCCGCCAGCGGATGCATGTCGTG 300



Db 81 GlycerLySGIntPrasnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGGCTATAGCAGCGCGCGGGCGGCGGAGCCCTCTATATGGCCAGGCGGATCTTGGGCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCCGCGCGCGCGCGCAAGCTCAAGCCTACCCCATGCGGAGCGCACCGGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGAGCGGACGCCATCTCGCGGTGCAGCATACGCCCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGlyGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTGTGTGCTGGAGGACACATCCAGCCATTGTGAAATACGCCCATGTACGCCCGGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCCACGTCGCGCTGCTGCGCGGAGCTTCAGCCTCTATCGCGGACATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGCAGGCGGCTGTCTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyGlyCysTyr 220  
QY 661 GTGCTGCGCGTGGTCGCGCGGACCGCTTTCGCGCGGAGATGATCAAGGTATTGTGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGGCGCGCGGCTTTCATGATTTTCGGGCGCGGACGCGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGGAGCGCGCTCCCGGAGAGCCGAAGAGGAGTGTGCTGCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTGCGCATGATCGCGTGTGGCCAAAGCGCGCGCGGATCCGGCGGCGCACTATTACGCGGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTACCGCGGCTGCTGCTGATCGAGCTCCGGCCCAACGCGCTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGCAACCGCAAAACGAGGACAGGAGCGCGCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
RESULT 8  
ADH35868  
ID ADH35868 standard; protein; 341 AA.  
XX  
AC ADH35868;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SegID34.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.  
XX  
OS unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.

XX (DIVE-) DIVERSA CORP.  
XX  
PA Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI Mcquaid J, Stege J;  
PI  
XX  
DR WPI; 2004-142708/14.  
DR N-PSDB; ADH35867.  
XX  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 34; 277pp; English.  
XX  
XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 341 AA;  
XX  
Alignment Scores:  
Pred. No.: 1.08e-118 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x ADH35868 (1-341)  
QY 1 ATGTGGAAGCCCATGACGAAGTATCGCGCGCGGTGACGCGCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATCTGACCGCACAGTCGAGAAAGCATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGACAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuTrpIleTrpLeu 60  
QY 181 GCGCGCGCGGCTTGGGCGATGCGCTTGTTCAGCGCTATTTCGAGAATTGCTGCGGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCCTGCGGATGCGCGCGCGCCGACGCGCATGATGTCGTG 300  
Db 81 GlycerLySGIntPrasnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGCTATAGCAGCGCGCGGCGGAGCCCTCTATATGGCCAGGCGGATCTTGGGCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCCGCGCGCGCGCGCAAGCTCAAGCCTACCCCATGCGGAGCGCACCGGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGAGCGGACGCCATCTCGCGGTGCAGCATACGCCCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGlyGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTGTGTGCTGGAGGACACATCCAGCCATTGTGAAATACGCCCATGTACGCCCGGAC 540

Db 161 AlAlauCyScySTrpGluHisIleGlnProleuSerLysTyraAlaMetTyraAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCTGCTGAGCCGAGCTTCAAGCTCTATCGCGGATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyraArgGlyMetAlaTyraAla 200  
QY 601 CTCGGACCCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyraAlaValGluGlyCysTy 220  
QY 661 GTGCTGGCTGCTGCGCGAGCCGTTTCGCGGAGATGATCAAGGTATTGGTGGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGCCGCGCGCGGTTCGCATGATTTTCGGAGCCGAGCGC 780  
Db 241 AspLysGluProleuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCCCGCTCCCGAGACCGAAGAGGAGTGTGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProleuProGluThrGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTGGCATGATCGCGTGTGGCCAAAGCCGCGCGCGATCCGCGCGGCCACTATTACGGCCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyraAlaArgPro 300  
QY 901 GACGTAACGGGCTGCTGCTGGATCGAGCTCCGGCCCAACGCGCTGTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGCAACCGCAAAAAGAGACAAGGCGAGCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 9  
ADG93569  
ID ADG93569 standard; protein; 341 AA.  
XX  
AC ADG93569;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE NitriIase enzyme amino acid sequence SegID34.  
XX  
KW nitriIase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003097810-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-090637/09.  
DR N-PSDB; ADG93568.  
XX  
PT New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitriIase activity, useful for screening enantioselective transformation.  
XX  
PS Claim 44; SEQ ID NO 34; 295pp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitriIase activity. NitriIase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated  
CC nitriIase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitriIase enzyme of the invention.

XX  
SQ Sequence 341 AA;

Alignment Scores:

Pred. No.:	1.08e-118	Length:	341
Score:	1464.00	Matches:	272
Percent Similarity:	90.1%	Conservative:	28
Best Local Similarity:	81.7%	Mismatches:	33
Query Match:	77.8%	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-1 (1-1041) x ADG93569 (1-341)

QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyraArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATTCGACCCGACAGTCGAGAAAGCATCGCGCTGATCGACAGCGCGCCAAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGGCCTGATCGCATTCGACAGACTTGATTCGCCGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTripleProGlyTyProleuTripleTyLeu 60  
QY 181 GCGCGCCGCGCTTGGGCGATGCGCTTCGTCAGCGCTATTTCAGAATTGCTGTCGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCTGCGCGGATGCGGCCCGCCGCAAGCATGCATGCTG 300  
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGCTATAGCAGAGCGCGCGCGCGAGCCCTTATATGGGCGAGGATCTTGGCGCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTTACCCTATGCGGAGCGCACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGAGACGGCAGCCATCTCGCGGTGCACAGTACCAGCATCGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrgLuthrGlyValGlyArgIleGly 160  
QY 481 GCGCTGTGCTGGAGCACATCCAGCCATGTGGAATAGCCCATGTACGCCCGCGAC 540  
Db 161 AlAlauCyScySTrpGluHisIleGlnProleuSerLysTyraAlaMetTyraAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCTGCTGAGCCGAGCTTCAAGCTCTATCGCGGATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyraArgGlyMetAlaTyraAla 200  
QY 601 CTCGGACCCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyraAlaValGluGlyCysTy 220  
QY 661 GTGCTGGCTGCTGCGCGAGCCGTTTCGCGGAGATGATCAAGGTATTGGTGGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGCGCGCGCGGTTCGCATGATTTTCGGAGCCGAGCGC 780  
Db 241 AspLysGluProleuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCCCGCTCCCGAGACCGAAGAGGAGTGTGTCGCCGATATCGAC 840

Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGGCATGATCGCGCTTGCGCCAAAGCGCGCCGATCCGGCGGCGCACTATTACGGGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTAACGCGCTGCTGCTGATCGACGTCGCCCAACGCGCTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGCAACCGCAAAACGAGACACAGCGCGCAGCGGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
RESULT 10  
AD162166  
ID AD162166 standard; protein; 341 AA.  
XX  
AC AD162166;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrlase polypeptide #17.  
XX  
KM Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipaeamic; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
DR WPI; 2004-090821/09.  
XX  
DR N-PSDB; AD162165.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrlase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 41; SEQ ID NO 34; 253pp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrlase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrlase polypeptide obtained from an  
CC environmental sample.  
XX  
SQ Sequence 341 AA;

Alignment Scores:  
Pred. No.: 1.08e-118 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x AD162166 (1-341)  
QY 1 ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGGTGCAGCGCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGACGAGCGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGlnLysThrIleAlaIleIleGluGlnAlaIleGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATTCGCGCTATCCCTTTGGATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCTTGCGGCGATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGTCGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCCCTGCGGATGCGGCCCGCCGACCGCATGATGTCGTG 300  
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGCTATAGCGAGCGCGCGCGCGGCGAGCCTCTATATGGGCCAGGCGCATTTCCGCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCCTACCATCGGACGACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGAGCGGACCATCTCGCGGTGCAGATACCGCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTTGCTGGGAGCACATCCAGCCATTGCGAATACGCCCATGTACGCCCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCTGTCGCGGAGCTTCAGCCTCTATCGCGGATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGGTCAATACCGCGGAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGCGCTGTCGCGGAGCCGTTTCGCCGAGATGATCAAGGTATTGTGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTCCGGCCCGACGGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGGAGCCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCGCGCATATGCAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGGCATGATCGCGCTTGCGCAAGCGCGCGCGCATCCGGCGGCGCACTATTACGGGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTAACGCGCTGCTGCTGATCGACGTCGCGGCCCAACGCGTGTCAAGCTTGATGCC 960



Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTCGAACCGCAAAACGAGACAAAGGCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
RESULT 11  
AD164287  
ID AD164287 standard; protein; 341 AA.  
XX  
AC AD164287;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrlase seq id 18.  
XX  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KW carboxylic acid; cyanohydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KW (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KW & diastomeric excess; food additive; drug intermediate; nitrlase.  
XX  
OS Unidentified.  
XX  
PN US2004014195-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 15-MAY-2003; 2003US-00440523.  
XX  
PR 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-121569/12.  
DR N-PSDB; AD164286.  
XX  
PT Novel isolated or recombinant polypeptide having nitrlase activity,  
PT useful in production of food additives.  
XX  
PS Claim 46; SEQ ID NO 34; 105bp; English.  
XX  
CC The invention describes an isolated or recombinant polypeptide (I)  
CC comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its  
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining & enantiomeric excess or & diastomeric  
CC excess. (I) is useful in the production of food additives and drug  
CC intermediates. This is the amino acid sequence of a nitrlase of the  
CC invention.  
XX  
SQ Sequence 341 AA;

Alignment Scores:  
Pred. No.: 1.08e-118 length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x AD164287 (1-341)  
QY 1 ATGTCGAGCCCATGACGAGATATCGCGCGCGGCTGACGCGCGGCTGTCCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATTCGACCGCACAGTCGAGAAACCGATCGCCCTGATCGAGCAGCGCCGCAAGCAGAC 120  
Db 21 AspleuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGACTTGATTCCCGGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrrIleProGlyTyrProLeuTrrIleTrrLeu 60  
QY 181 GCGCGCGCGCTTGCGCATGCGCTTCGTCAGCGCTATTCGAAATTCGAAATTCGCTGCGC 240  
Db 61 GlySerProAlaTrrPglYmeThrArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCGCTGGCGGATGCGCGCGCGCCAGCGCATGCATGCTG 300  
Db 81 GlySerLysGlnTrrPasnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGCTATAGCGAGCGCGCGCGCGGCGAGCCTCTATATGGGCCAGCGCATCTTCGGCCC 360  
Db 101 ValGlyPheSerGlnTrrPasnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 120  
QY 361 GATGCGCATTCGATCGCGCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGGAGGAGAGACGGGAGCCATCTCGCGTGACGATACGCCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTTGCTGGAGACATCCAGCCATGTGCAATTCAGCCATGTACGCCGCGCAC 540  
Db 161 AlaLeuCysCysTrrPgluHisIleGlnProLeuSerLysTrrAlaMetTrrAlaAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCGCTGTCGCCGAGCTTCAGCCTTATCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrrProCysPheSerLeuTrrArgGlyMetAlaTrrAla 200  
QY 601 CTCGGACCGGAGGTCAATACCGCGCAAGCCAGATCTACGGGTGAGGGCGGCTGCTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTrrAlaValGluGlyCysTrr 220  
QY 661 GTGTCGCGCTGTCGCGCGCGCTTCGCCGAGATGATCAAGGTATTGTTGATACGCCCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAGGAGATGTTCTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGCCGACGCGC 780  
Db 241 AspyrGluProLeuLeuLeuAlaGlyGlyLysPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCTTGCGCGAGCGCGCTCCCGAGACCGAAGAGAGGAGTCTGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGCGATGATCGCGTTGGCCAAAGCGCGCGCGCGATCCGCGCGGCCACTATTCAGGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTrrAlaArgPro 300  
QY 901 GACGTAAAGCGCGCTGCTGCTGATCGATCGCGGCCCAAGCGGTCTGACGCTTGATGCC 960



Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGGAACCGCAAAACGAGACAGCGCGCGCCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 12  
ADCC23749

ID ADC23749 standard; protein; 348 AA.

AC ADC23749;

DT 18-DEC-2003 (first entry)

DE Protein sequence (SeqID 16) exhibiting nitrilase activity.

KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.

OS Unidentified.

PN WO2003000840-A2.

PD 03-JAN-2003.

PF 15-MAY-2002; 2002WO-US015983.

PR 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309006P.

PR 22-JAN-2002; 2002US-0351336P.

XX (DIVE-) DIVERSA CORP.

PA (MADD/) MADDEN D.

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;

XX WPI; 2003-201417/19.

DR N-PSDB; ADC23748.

PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.

XX Claim 40; SEQ ID NO 16; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.

XX SQ Sequence 348 AA;

Alignment Scores:

Pred. No.: 1.97e-96 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 7 Gaps: 2

US-09-751-299-1 (1-1041) x ADC23749 (1-348)

QY 10 CCCATGACGAAG--TATCGCGCGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGACGTGCGC 126  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaIleSerAsnGlyAlaLys 41  
QY 127 CTGATCGCATTCACAGACTTGATTCACGCTATTCGAGAAATTCGCTCGTGCAGCAGC 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTCGAGAAATTCGCTCGTGCAGCAGC 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGGCGCCTGGCGGATGGCGCGCGCGCGCCAGCGCATGATGTCGTGCGCGC 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101  
QY 307 TATAGCGAGCGCGCGCGCGCGCTCTATATGCGGCCAGCGCATCTTCGCGCCGATGCGC 366  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCGCTCAAGCCTACCACTGCGGAGCGCAGCCGCTTCGCGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGCGCAGCCCATCTCGCGGTGCACAGATACCGGCATCGGCGCGCTCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCysValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTGTGGGAGACACATCCAGCCATTTGTGGAATACGCCATGTACGCCCGCGCAGCAAG 546  
Db 162 CysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGln 181  
QY 547 GTCCACGTGCGGTGCGCGCGCGCTTCAGCCTCTATTCGCGCGCATGGCCTATGCGCTTCGGA 606  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGCAGGCGCGCTGCTAGTGTCTG 666  
Db 202 ProGluLeuAsnAsnAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTGTGCGCGCAGCGCTTTCGCGGAGATGATCAAGTATTGTGTGATACGCCGCAAG 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGCGCGCGCGGTTCGCTGATTTTCGGGCGCGCGCGCGCGC 786  
Db 242 GluProLeuLeuGluGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGGCCGAGCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCGCGATATCGACTTCGCGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGGTTGGCAGAGCGCGCGCGCTGATTCGCGCGCGCGCGCGCGCGCTA 906  
Db 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301  
QY 907 ACGCGGCTGTGTCGATCGACGTCCGCGCCCAAGCGCTGTACGCTTGATTCGCGCATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrTrpLeuAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGACAAGGCGCAGCGCGCGCTGCGCGTGTGCGGCAAAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGlnGln 340  
QY 1027 GCCGCGCGCG 1035  
Db 341 LysAlaAla 343

RESULT 13

```

ADH35850
ID ADH35850 standard; protein; 348 AA.
XX
XX ADH35850;
AC
DT 11-MAR-2004 (first entry)
DE
XX
XX Chemical process monitoring-related nitrilase protein sequence SegID16.
KM chemical process monitoring; biochemical process monitoring; cyanide;
KW high throughput system; enzyme.
XX
XX Unidentified.
OS
XX
XX WO2003098187-A2.
XX
XX 27-NOV-2003.
XX
XX 15-MAY-2003; 2003WO-US015639.
XX
XX 15-MAY-2002; 2002US-0380737P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI Mcquaid J, Stege J;
PI
DR WPI; 2004-142708/14.
XX
XX N-PSDB; ADH35849.
XX
XX
XX Monitoring a chemical or biochemical process comprises providing a
PT reactant comprising a cyanide or a material that can be converted to
PT cyanide or a reactant that generates a cyanide or a material that can be
PT converted to cyanide.
XX
XX Claim 73; SEQ ID NO 16; 277pp; English.
XX
XX This invention relates to a novel method of monitoring chemical or
CC biochemical processes. The method involves providing a reactant
CC comprising cyanide (or a material that can be converted to a cyanide)
CC that generates as a reaction product cyanide or a material that can be
CC converted to cyanide and measuring the concentration of produced cyanide.
CC The method is useful for monitoring a chemical or biochemical process.
CC The method is effective for high throughput systems and is sufficiently
CC sensitive to detect a small amount of product. The present sequence is
CC that of a nitrilase protein which may be used in the method of the
CC invention.
XX
XX
XX Sequence: 348 AA;
SQ
Alignment Scores:
Pred. No.: 1.97e-96 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conservative: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Indels: 2
DB: 8 Gaps: 2
US-09-751-299-1 (1-1041) x ADH35850 (1-348)
QY 10 CCCATGACGAAG--TATCGCGGCGGCGGTGCAGGCGCGCGGTGTTCTGATCTC 66
||| ::||| ::||| ||||| ||||| ||||| ||||| |||||
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21
QY 67 GACCGACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGACGTGCGC 126
::: :::: ||||| ||||| ||||| ||||| ||||| |||||
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGlyAlaAlaSerAsnGlyAlaLys 41
QY 127 CTGATCGCATTCACAGAGACTTGATTCGCCGCTATCCCTTTTGATATGGCTGGCGCG 186
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 42 LeuIleAlaPheProGluThrTriPleProGlyTyrProTriPleTriPleAspSer 61
QY 187 CCGGCTTGGGGCATGCGCTTCGTCACGCGCTATTTCGAGAATTGCTCGCGCGCACC 246

```

Db	62	ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer	81
QY	247	AAGCAGTGGCAGGCGCTGGCGGATGCGGCCCGCCGACGGCATGCTGTCGGCGGC	306
Db	82	GlulGlnAlaLysArgMetAsnGlnAlaAlaAsnAsnLysIleTyrValValMetGly	101
QY	307	TATAGCAGCGCGCGCGCGGAGCCTTATATAGGCGCAGCGCATCTTCGCCCCGATGC	366
Db	102	TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly	121
QY	367	GATCTGATCGCGCGCGCGCGGCAAGCTCAAGCCTACCACATGCCAGCGCACCGTGTCGC	426
Db	122	GlutThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly	141
QY	427	GAGGAGACCGGACGACCATCTCGCGGTGCACAGATACCGGCATCGGCGCCTCGCGCGCTC	486
Db	142	GlulGlyAspGlySerHisLysLeuCysValMetAspThrGlnIleGlyArgValGlyAlaMet	161
QY	487	TGTTGCTGGAGACATCCAGCCATTGTGGAATAGCCCATGTACGCCGCGCAGCAAG	546
Db	162	CysCysTrpGluHisIleLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGlyGln	181
QY	547	GTCCACGTCGCGTCTGTGGCGGAGCTTCAAGCCTCTATCCGCGCATGGCCTATGCGTCCGA	606
Db	182	IleHisIleAlaSerTyrProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly	201
QY	607	CCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTACGTGCTG	666
Db	202	ProGluLeuAsnAsnAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu	221
QY	667	GCGTCGTGCGGACCGGTTTCGCGGAGATGATCAAGGATTTGGGGCCCGACGCGCGGCC	726
Db	222	AlaProCysAlaThrValSerLysGlnMetIleGlnMetLeuIleAspAspProAlaGlyLys	241
QY	727	GAGATGTTCTCAAGCGCGCGCGGCGGTTTGGCCATGATTTTCGGGCGCGACGCGCGGCC	786
Db	242	GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro	261
QY	787	CTGGCCGAGCGGCTCCCGGAGAGCCGAGAGGAGTGTGTGCGCGGATATGACCTCGGC	846
Db	262	LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGly	281
QY	847	ATGATCGCGTTGGCCAGGCGGCGCGCGCATCCGCGCGGACCACTATTCACGCGCGCGCTA	906
Db	282	MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlnHisTyrAlaArgProAspVal	301
QY	907	ACGCGGCTGCTGTGATCGACGTCCGCGCCCAACGCGTCTCACGCTTGATGCCGCAATC	966
Db	302	ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer	321
QY	967	GAACCGCAAAACGAGAGCAAGGGCGACGCGCGCGCTGCGGTGTGGCGGAAGCGCC	1026
Db	322	GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGlnGln	340
QY	1027	GCCGCGCGC 1035	
Db	341	LysAlaAla 343	
RESULT 14			
ID	ADG93551	standard; protein; 348 AA.	
XX	ADG93551;		
DT	11-MAR-2004	(first entry)	
DE	NitriIase enzyme amino acid sequence SegID16.		
XX	nitriIase; nitriIe; carboxylic acid; chemical process; pH; temperature;		
KW	enantioselective transformation; enzyme.		
XX			
OS	Unidentified.		

XX WO2003097810-A2.  
XX 27-NOV-2003.  
XX 15-MAY-2003; 2003WO-US015712.  
XX 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
XX MPI; 2004-090637/09.  
DR N-PSDB; ADG93550.  
XX  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
XX Claim 44; SEQ ID NO 16; 295bp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.  
XX  
SQ Sequence 348 AA;

Alignment Scores:  
Pred. No.: 1.97e-96 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 8 Gaps: 2

US-09-751-299-1 (1-1041) x ADG93551 (1-348)

QY 10 CCCATGACGAAG--TATCGCGGCGCGGTGACGCGCGCGGTGTTCTCATCTC 66  
||| :||| :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCGAGAAAGCGATCGCGCTGATCGAGCAGCGGCCAAGCAGCAGCTGCGC 126  
::: ::| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaAlaSerAsnGlyAlaLys 41  
QY 127 CTGATCGCATTCCAGAGACTTGATTTCCCGGCTATCCCTTTTGATATGCTGCGCGC 186  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGCGCATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTCGTGCGCGCAGC 246  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGGCGCTGGCGGATGCGGCCCGCCGACGCGCATGATGTCGTGCGCGC 306  
::: ||| ::| ::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101  
QY 307 TATAGCGAGCGCGCGCGCGGCGGAGCTTATATGGGCCAAGCGATCTTCGGCCCGATGGC 366  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCCGCGCGCGCGCGCAAGCTCAAGCCTACCATGCGGACGCGCACCGTGTTCGGC 426  
::: ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGGGACGACCATCTCGCGGTGACAGATACCGCCCATCGGCGCGCTC 486

Db 142 GluGlyAspGlySerHisLeuCyValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTTCGGGAGACACATCCAGCCATTGTGAAATACGCCCATGTACCGCCGACGAACAG 546  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 162 CysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetCysSerGlnAspGluGln 181  
QY 547 GTCCACGTGCGCTGCTGTGCGCGGAGCTTCAAGCTCTATCGCGGCGATGCGCTCGGA 606  
::: ||| ::| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGGAGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGCGCTGCTACGTGCTG 666  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 202 ProGluLeuAsnAsnAlaAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCGTGCGCGACCGCTTTCGCGCGAGATGATCAAGTATTGGTGATACGCCGACAAG 726  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGGCGCGCGCGGCTTTTGCCATGATTTTCGGGCCGACGCGCGCC 786  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 242 GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGCGCGGACCGGCTCCCGGAGACCCGACAGGAGACTGCTGTGCGCGCATATCGACCTCGC 846  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 262 LeuAlaLysProLeuProGluAsnGluGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGCGCTTGCGCCCAAGGCGCGCGCGCATCCGCGGGCGCCACTATTCACGCGCCGACGTA 906  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301  
QY 907 ACGCGGCTGCTGCTGATCGACGTCGCGGCCCAACGCGCTGTCACGCTTGATGCGCATTC 966  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGGACCAAGGGCGACGCGCGCGCGCTGCGCGCTGCGCGGCAAAAGCGCC 1026  
||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGluMetGluAlaGluGln 340  
QY 1027 GCCGCGCGCG 1035  
||| |||||  
Db 341 LysAlaAla 343

RESULT 15  
ADI62148  
ID ADI62148 standard; protein; 348 AA.  
XX  
AC ADI62148;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase polypeptide #8.  
XX  
KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipemic; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX

PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX WPI; 2004-090821/09.  
DR N-PSDB; ADI62147.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 41; SEQ ID NO 16; 253bp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrilase polypeptide obtained from an  
CC environmental sample.  
XX  
SQ Sequence 348 AA;

Alignment Scores:  
Pred. No.: 1.97e-96 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 8 Gaps: 2

US-09-751-299-1 (1-1041) x ADI62148 (1-348)

QY 10 CCCATGACGAG--TATCGCGCGCGCGGTGCAGCGCGCGCGGTCTTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCGAGAAAGCGATCGGCTGATCGACGAGCGCGCCACAGGACGATGCGC 126  
Db 22 GluGluAlaIleSerLysGlyIleSerLeuIleGluGluAlaAlaSerAsnGlyAlaLys 41  
QY 127 CTGATCGCATTTCCAGAGACTTGATTCGCCGCTATCCCTTTTGATGATGCTGGCGCG 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGCGATGCGCTTCCTCCAGCGCTATTTCGAGATTCCGTCGTGCGCGCAGC 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGGCGCTGCGGATGCGCGCGCGCGCCACGCGCATGCTGCTGCGCGGC 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101  
QY 307 TATAGCAGCGCGCGCGCGCGCGCTTATATGCGCGAGCGCATCTTCCGCCGATGCG 366  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGAGCGCAACGTTGTCGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuValProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGCGCAGCATCTCGCGTGCAAGTACCGCCATCGGCGCGCTCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisIleuCyValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTGTGCTGGAGACATCCAGCCATTTCGAATAAGCCATGTACCGCGCGCGAGACAG 546

Db 162 CysCysTrpGluHisIleuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGln 181  
QY 547 GTCCACGTCGCGCTCGTGCGCGAGCTTACGCTCTATCCCGGATGCGCTCGCA 606  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGGAGTCATATACCGCGCGCAAGCAGATCTACGCGGTGAGGCGGCTGCTACGTGCTG 666  
Db 202 ProGluLeuAsnAsnAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCGTGCGCGACCGTTTCGCGCGAGATGATCAAGTATTGTTGATACGCCGACAAG 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGCGCGCGCGGTTTTCGATGATTTTCGCGCGCGCGCGCGCC 786  
Db 242 GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGCGCGACCGCGCTCCCGGAGACCGAAGAGGAGCTGCTGTCGCCGATATGACCTCGGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGCGTTGCGCAAGCGCGCGCGATCCGCGCGCGCACTATTCACGCGCGCGAGTA 906  
Db 282 MetIleSerMetAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301  
QY 907 ACGCGCTGCTGCTGATGACGTCGCGCGCGCGCGCGCGCTGCTGATGCGCGCATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGACAGGAGCGCGCGCGCGCGCGCTGCGGTGCGGAAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGln 340  
QY 1027 GCCGCGCGCG 1035  
Db 341 LysAlaAla 343

Search completed: April 27, 2006, 01:18:36  
Job time : 181.007 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:35:46 ; Search time 7.2946 Seconds  
(without alignments)  
2359.702 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1881

Sequence: 1 atgtcgagagcccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 segs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090039\_18719/app\_query.fasta\_1  
-DB=Issued\_Patents\_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOFC=0  
-LOOEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs02p -USER=US09751299 @CGN 1 1 101 @runat\_26042006\_090039\_18719  
-NCPU=6 -ICPU=3 -NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	940	50.0	344	1	US-08-690-493-1 Sequence 1, Appli
2	869	46.2	356	2	US-09-806-876A-2 Sequence 2, Appli
3	807	42.9	369	2	US-09-823-373-5 Sequence 5, Appli
4	807	42.9	369	2	US-09-823-373-14 Sequence 14, Appli
5	592.5	31.5	354	1	US-08-447-702-5 Sequence 5, Appli
6	592.5	31.5	354	1	US-08-465-615-5 Sequence 5, Appli
7	255.5	13.6	863	2	US-09-252-991A-26099 Sequence 26099, A
8	247.5	13.2	467	2	US-09-252-991A-18296 Sequence 18296, A
9	244.5	13.0	419	2	US-09-252-991A-31734 Sequence 31734, A
10	242.5	12.9	724	2	US-09-252-991A-31715 Sequence 31715, A
11	242.5	12.9	757	2	US-09-252-991A-23569 Sequence 23569, A
12	239	12.7	369	2	US-09-252-991A-25394 Sequence 25394, A

C 13	238.5	12.7	344	2	US-09-252-991A-22486	Sequence 22486, A
C 14	238.5	12.7	681	2	US-09-252-991A-24567	Sequence 24567, A
C 15	237	12.6	977	2	US-09-252-991A-16655	Sequence 16655, A
C 16	236.5	12.6	539	2	US-09-252-991A-17070	Sequence 17070, A
C 17	236.5	12.6	618	2	US-09-252-991A-28358	Sequence 28358, A
C 18	236.5	12.6	759	2	US-09-252-991A-29395	Sequence 29395, A
C 19	235.5	12.5	1008	2	US-09-252-991A-29419	Sequence 29419, A
C 20	235	12.5	618	2	US-09-252-991A-23696	Sequence 23696, A
C 21	235	12.5	720	2	US-09-252-991A-21881	Sequence 21881, A
C 22	233.5	12.4	882	2	US-09-413-814-78	Sequence 78, Appl
C 23	232.5	12.4	335	2	US-09-252-991A-23674	Sequence 23674, A
C 24	232.5	12.4	836	2	US-09-252-991A-26065	Sequence 26065, A
C 25	231	12.3	552	2	US-09-252-991A-18137	Sequence 18137, A
C 26	229	12.2	638	2	US-09-252-991A-27068	Sequence 27068, A
C 27	229	12.2	757	2	US-09-252-991A-23569	Sequence 23569, A
C 28	227.5	12.1	611	2	US-09-252-991A-17597	Sequence 17597, A
C 29	227.5	12.1	722	2	US-09-252-991A-24102	Sequence 24102, A
C 30	226.5	12.0	511	2	US-09-252-991A-26078	Sequence 26078, A
C 31	226.5	12.0	541	2	US-09-252-991A-17206	Sequence 17206, A
C 32	226	12.0	525	2	US-09-252-991A-23981	Sequence 23981, A
C 33	225.5	12.0	438	2	US-09-252-991A-32407	Sequence 32407, A
C 34	225.5	12.0	1706	2	US-09-252-991A-31760	Sequence 31760, A
C 35	225	12.0	439	2	US-09-252-991A-17127	Sequence 17127, A
C 36	225	12.0	631	2	US-09-252-991A-30909	Sequence 30909, A
C 37	223.5	11.9	257	2	US-09-252-991A-31869	Sequence 31869, A
C 38	223.5	11.9	421	2	US-09-252-991A-22123	Sequence 22123, A
C 39	223.5	11.9	518	2	US-09-252-991A-23604	Sequence 23604, A
C 40	223.5	11.9	561	2	US-09-252-991A-23080	Sequence 23080, A
C 41	223.5	11.9	1213	2	US-09-413-814-79	Sequence 79, Appl
C 42	223	11.9	720	2	US-09-252-991A-21881	Sequence 21881, A
C 43	222.5	11.8	366	2	US-09-252-991A-22528	Sequence 22528, A
C 44	222.5	11.8	491	2	US-09-252-991A-31758	Sequence 31758, A
C 45	222.5	11.8	718	2	US-09-252-991A-32743	Sequence 32743, A

ALIGNMENTS

RESULT 1  
US-08-690-493-1  
; Sequence 1, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: YU, Fujio  
; TITLE OF INVENTION: No. 5872000e1 Nitrlase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acid residues

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gordona terrae  
STRAIN: MA-1  
CELL TYPE: unicellular organism  
FEATURE:  
OTHER INFORMATION: Xaa is Met or a deletion  
US-08-690-493-1

Alignment Scores:  
Pred. No.: 3.32e-73 Length: 344  
Score: 940.00 Matches: 178  
Percent Similarity: 69.7% Conservative: 54  
Best Local Similarity: 53.5% Mismatches: 98  
Query Match: 50.0% Indels: 3  
DB: 1 Gaps: 1

US-09-751-299-1 (1-1041) x US-08-690-493-1 (1-344)

QY 22 TATCGCGCGCGCGGTGCAGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAG 81  
Db 9 PheLysAlaIaValThrGlnAlaGluProValTrrPheAspLeuSerAlaThrValAsp 28  
QY 82 AAAGCGATCGGCTGATCGAGCAGCGCGCGCACAGAGCAGTGGCTGATCGCATTCCTCA 141  
Db 29 LysThrIleAlaLeuValGluGlnAlaSerAlaGlyAlaAspLeuIleAlaPhePro 48  
QY 142 GAGACTTGGATTCCCGGCTATCCCTTTGGATATGCTGGCGCGCGCTTGGGGCATG 201  
Db 49 GluThrTrrPleProGlyTrrProTrrPheLeuTrrPleuAspSerAlaIaTrrGlnSer 68  
QY 202 CGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGCGCGCGCACAGCAGTGGCAGGCC 261  
Db 69 GlnTrrPheIleArgTrrProGlnAsnSerLeuAspLeuAspGlySerGluPheAlaIa 88  
QY 262 CTGGCGATCGCGCGCGCGCGCGCGCGCATGCTGCGCGCGCTATAGCAGCGCGCG 321  
Db 89 IleArgGluAlaIaIaArgLysAsnAspIleAlaIleThrMetGlyPheSerGluArgGly 108  
QY 322 GCGCGCAGCTCTATATGGCCAGCGCATCTTCGCGCGCGCTGCTGCTGCGAGCAC 381  
Db 109 HisGlySerLeuTrrMetGlyGlnAlaValIleGluArgAspGlyValValArgThr 128  
QY 382 CGCCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTGTTCGCGAGGAGAGCGCAGC 441  
Db 129 ArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGluGlyAspGlySer 148  
QY 442 CATCTCGGGGTGCACGATACCGCCATCGGGCGCGCTCGCGCGCTGTGCTGGAGCAC 501  
Db 149 AspLeuValValAspGlnThrSerLeuGlyArgValGlySerLeuCysCysTrrGlnHis 168  
QY 502 ATCCAGCATTTGTGGAATACGCCATGTACCGCGCGCGCACGACAGGTCCACGTCCGCTCG 561  
Db 169 LeuGlnProLeuThrLysTrrAlaMetTrrSerGlnHisGluGlnIleHisIleAlaIa 188  
QY 562 TGGCCAGCTTCAGCCTCTATCGCGCGCATGGCTATGCGCTCGGACCGGAGGTCAATACC 621  
Db 189 TrrProSerPheSerIlePheProGlyAlaValTrrAlaLeuGlyProGluValAsnThr 208  
QY 622 GCGCGCAAGCAGATCTACGCGGTGAGGGCGCGCTGCTACGCTGCGCTGCGCGCAC 681  
Db 209 AlaAlaSerGlnGlnTrrAlaValGluGlyGlnThrTrrValLeuAlaProCysAlaVal 228  
QY 682 GTTTCGCGCGAGATGATCAAGGTATGTGGATACGCCGACAGAGAGATGTTCTCAAG 741  
Db 229 IleGlyAspAlaGlyTrrGluAlaPheAlaAspThrGluGluLysArgGlnLeuIleHis 248  
QY 742 GCGCGCGCGGTTTGGCATGATTTTCGGCGCGCGCGCGCGCGCTGCGCGAGCGGCTC 801  
Db 249 LysGlyGlyGlyTrrAlaArgIleTrrGlyProAspGlyArgSerLeuAlaGluProLeu 268

QY 802 CCGGAGACGAGAGGAGCTGCTGCTGCCGATATCGACCTCGCGCATGTCGCTTGGCC 861  
Db 269 AlaProAsnAspGluGlyIleLeuTrrAlaAspIleAspLeuSerAlaIleLeuAlaIa 288  
QY 862 AAGCGCGCGCGATCCGCGCGCGCATTAATTCAGCGCGCGCGCATGAGCGGCTGCTG 921  
Db 289 LysAsnProAlaAspProValGlyHisTrrSerArgProAspValLeuArgLeuGlyPhe 308  
QY 922 GATCAGCTCCGGCCCAACGCGTGTCAAGCTTGATGCCGCAATTCGAACCGCAAAACGAG 981  
Db 309 AsnLysAlaProGlnProLysValAsnIleLeuGlyThr-GluProSerArgThrThre 328  
QY 982 GACA-----AGGCGACGCGCGCGCGCTGCGCGTGG 1012  
Db 328 rThrGlnCysArgProThrThrIleArgArgSerTrr 340

RESULT 2

US-09-806-876A-2  
Sequence 2, Application US/09806876A  
Patent No. 6869783

GENERAL INFORMATION:

APPLICANT: Resz-Loeschke, Marion  
APPLICANT: Friedrich, Thomas  
APPLICANT: Hauer, Bernhard  
TITLE OF INVENTION: A process for preparing chiral carboxylic acids  
TITLE OF INVENTION: from nitriles using a nitrilase or microorganisms  
FILE REFERENCE: 49462  
CURRENT APPLICATION NUMBER: US/09/806, 876A  
CURRENT FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: Germany/19848129.2  
PRIOR FILING DATE: 1998-10-19  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: wordPerfect version 6.1  
SEQ ID NO 2  
LENGTH: 356  
TYPE: PRT  
ORGANISM: Alcaligenes faecalis  
US-09-806-876A-2

Alignment Scores:

Pred. No.: 4.87e-67 Length: 356  
Score: 869.00 Matches: 171  
Percent Similarity: 66.2% Conservative: 54  
Best Local Similarity: 50.3% Mismatches: 105  
Query Match: 46.2% Indels: 10  
DB: 2 Gaps: 2

US-09-751-299-1 (1-1041) x US-09-806-876A-2 (1-356)

QY 25 CCGCGCGCGCGGTGCAGCGCGCGCGGTGTTCTCGATCTCGACCGCACAGTCGAGAAA 84  
Db 8 ArgAlaAlaAlaValGlnAlaIaIaSerProAsnTrrAspLeuAlaThrGlyValAspLys 27  
QY 85 GCGATCGGCTGATCGAGCAGCGCGCGCACAGCAGTGGCTGATCGCATTCCTCAAG 144  
Db 28 ThrIleGluLeuAlaArgGlnAlaIaArgAspGluGlyCysAspLeuIleValPheGlyGlu 47  
QY 145 ACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGCGCGCGCTTGGGCGATGCGC 204  
Db 48 ThrTrrPleuProGlyTrrProPheHisValTrrPleuGlyAlaProAlaTrrPheSerLeuLys 67  
QY 205 TTCGTCACGCGTATTTGAGAATTGCTGCGCGCGCGCACAGCAGTGGAGGCGCTG 264  
Db 68 TrrSerAlaArgTrrTrrAlaAsnSerLeuSerLeuAspSerAlaGluPheGlnArgIle 87  
QY 265 GCGGATGCGCGCGCGCGCGCGCATGATGCTGCGCGCGCGCTATAGCGAGCGCGCGGC 324  
Db 88 AlaGlnAlaIaIaArgThrLeuGlyIlePheIleAlaLeuGlyTrrSerGlnArgSerGly 107  
QY 325 GCGAGCTCTATATGAGCGCAGCGCATCTTCGCGCGCGCGCGCGATGCTGATCGCGCGCGC 384  
Db 108 GlySerLeuTrrLeuGlyGlnCysLysLeuIleAspAspLysGlyGluMetLeuTrrPheArg 127





Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIaGluGlnGlu 344  
QY 1027 GCCCGCCGCG 1035  
Db 345 ArgGlnAla 347  
RESULT 4  
US-09-823-373-14  
; Sequence 14, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: Dicosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; TITLE OF INVENTION: Acidovorax Facilis 72W  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 14  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Acidovorax Facilis  
US-09-823-373-14  
Alignment Scores:  
Pred. No.: 1.18e-61 Length: 369  
Score: 807.00 Matches: 160  
Percent Similarity: 61.8% Conservative: 52  
Best Local Similarity: 46.6% Mismatches: 127  
Query Match: 42.9% Indels: 4  
DB: 2 Gaps: 2  
US-09-751-299-1 (1-1041) x US-09-823-373-14 (1-369)  
QY 16 ACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTCGATCTCGACCGCACA 75  
Db 6 SerLysPheLeuAlaAlaThrValGlnIaGluProValTyrLeuAspAlaAspAlaThr 25  
QY 76 GTCGAGAAAGCGATCGCGCTGATCGAGCAGCGCGCCACAGCAGCAGCGCTGATCGCA 135  
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaIaGlnLysGlyAlaSerLeuIleAla 45  
QY 136 TTCCCAAGACTTGATTCGCCGCTATCCCTTTTGATATGGCTGGCGCGCGCTTGG 195  
Db 46 PheProGluValPheIleProGlyTyrProTyrTrrPalatTrrLeuGlyAspValLysTyr 65  
QY 196 GGCATGCGCTTCGTCACGCGCTATTCGAGAATTGCTCGTGGCGCGCAGCAGCAGTGG 255  
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85  
QY 256 CAGGCCCTGGCGGATGCGCGCGCGCGCCACGCGCATGCATGCTGGCGCGCTATAGCGAG 315  
Db 86 ArgArgLeuGlnLeuAlaIaArgAlaGlnLysIleAlaLeuValMetGlyTyrSerGlu 105  
QY 316 CGCGCGCGCGCAGCCTCTATATGGGCCACGGCGATCTTCGGCCCCGATGGCGATGATC 375  
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
QY 376 GCCGCGCGCGCAAGCTCAAGCCTAACCCATGCGAGCGCACCGTGTTCGGCAGGAGAC 435  
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145  
QY 436 GGCAGCCATCTCGCGGTGACAGCATACCGCATCGGCGCGCTCGCGCGCTGTGTGCTGG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCySerTrp 165

QY 496 GAGCATCCAGCCATTGTCGAAATACGCCATGTACGCCCGCCGACGAAAGTCCACGTC 555  
Db 166 GluHisPheGlnProLeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisVal 185  
QY 556 GCGTCGTGGCCGAGCTTCAGCCTCTATCGCGCATGGCCTATGCGCTCGACCGGAGTC 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205  
QY 616 AATACCGCCGCAAGCCAGATCTACGCGGTGACGAGCGCGGTGCTACGTGCGTGTGC 675  
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCySerThr 225  
QY 676 GCGACCGTTTCGCCCGAGATGATCAAGTATTGTTGATACGCCCGACAGAGATGTTTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCybLeuAsnAspGluGlnArgAlaLeu 245  
QY 736 CTCAGGCGCGCGCGCGGTTCCTCATGATTTTCGGGCGCGCGCGCGCGCGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTrrPalaaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
QY 796 CCGCTCCCGGAGACCGAAGAGGACTGCTGTGCGCCGATATCGACCTCGGCATGATCGCG 855  
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285  
QY 856 TTGCGCCAAAGCGCGCGCGCGATCCGCGCGCCACTATTCACGCGCGCATCGCGCTG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305  
QY 916 CTGCTGATCGACGT-----CCGCGCCACGCGTGTACGCTTGATGCCGATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324  
QY 967 GAACCGCAAAACGAGCAGAGGCGCGCGCGCGCGCTGCGGTGTCGCGAAAGCGCC 1026  
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIaGluGlnGlu 344  
QY 1027 GCCCGCCGCG 1035  
Db 345 ArgGlnAla 347  
RESULT 5  
US-08-447-702-5  
; Sequence 5, Application US/08447702  
; Patent No. 5629190  
; GENERAL INFORMATION:  
; APPLICANT: Petre, Dominique  
; APPLICANT: Cerbeleaud, Edith  
; APPLICANT: Levy-Schil, Sophie  
; APPLICANT: Crouzet, Joel  
; TITLE OF INVENTION: Polypeptides Possessing A Nitrilase  
; TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,  
; Patent No. 5629190  
; TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them  
; TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy-  
; TITLE OF INVENTION: lates by Means of Said Polypeptides  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/447,702  
; FILING DATE: 23-MAY-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92-09-882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-702-5

Alignment Scores:  
Pred. No.: 4.76e-43 Length: 354  
Score: 592.50 Matches: 137  
Percent Similarity: 54.0% Conservative: 51  
Best Local Similarity: 39.4% Mismatches: 133  
Query Match: 31.5% Indels: 27  
DB: 1 Gaps: 7

US-09-751-299-1 (1-1041) x US-08-447-702-5 (1-354)

QY 31 GCGGCGGTGACAGCGCCGCGGTGTTCTGATCTCGACCGCACAGTCAAGAAAGCATC 90  
Db 10 AlaAlaValGlnAlaAlaProValPheMetAsnLeuGlnAlaThrValAspLysThrCys 29  
QY 91 GGCTGTATCGACAGCGCGCCCAAGCAGACGCTGCTGATCGCATTCGACAGACTTG 150  
Db 30 LysLeuIleAlaGlnAlaAlaLaserMetGlyAlaLysValIleGlyPheProGlnAlaPhe 49  
QY 151 ATTCCGCGCTATCCCTTTTGATATGGCTGGC-----GCGCCGCTTG 195  
Db 50 IleProGlyIleProTyrTrpIleTrpThrSerAsnMetAspPheThrGlyMetMetTrp 69  
QY 196 GGCATGCGCTTCCTCCAGCGCTATTTCGAATTCGCTCGTCCGCGGACCAAGCAGTGG 255  
Db 70 AlaVal-----LeuPheLysAsnAlaIleGluIleProSerLysGluVal 84  
QY 256 CAGGCGCTGCGGATCGGCGCCGCGCAGCGCATGTCGTGCGCGCTATAGCGAG 315  
Db 85 GlnGlnIleSerAspAlaAlaLysLysAsnGlyValTyrValCysValSerValSerGlu 104  
QY 316 CCGCGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGCGCCGCGCATGCGATCTGATC 375  
Db 105 LysAspAsnAlaSerLeuTyrLeuThrGlnLeuTyrPheAspProAsnGlyAsnLeuIle 124  
QY 376 GCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTTCGCGGAGGAGAC 435  
Db 125 GlyLysHisArgLysPheLysProThrSerSerGluArgAlaValTrpGlyAspGlyAsp 144  
QY 436 GGCAGCCATCTGCGGTCACGATACCGGCATCGGGCGCTCGGCGCGCTCTGTGCTGG 495  
Db 145 GlySerMetAlaProValPheLysThrGlnTyrGlyAsnLeuGlyGlyLeuGlnCysTrp 164  
QY 496 GAGCACATCCAGCCATTGTGAATATCGCATGTACCGCGCGGACGACAGAGTCCACGTC 555  
Db 165 GluHisAlaLeuProLeuAsnIleAlaIleMetGlySerLeuAsnGlnGlnValHisVal 184  
QY 556 GCGTCGTGGCGAGCTTC-----AGCCTTATCGCGGATGCGC 594  
Db 185 AlaSerTrpProAlaPheValProLysGlyAlaValSerSerArgValSerSerVal 204  
QY 595 TATGCGCTCGGACCGGAGGTCAATACCGCGCAGCGCATCTACGCGGTGAGGCGCGC 654  
Db 205 CysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyrAlaIleSerAsnGln 224

QY 655 TGCTACGTGCTGGCGTGTGCGCGCACCGCTTTCGCCGAGATGATCAAGTATTGTGAT 714  
Db 225 ValTyrValIleMetSerThrAsnLeuValGlyGlnAspMetIleAspMetIleGlyLys 244  
QY 715 ACGCCCGCAAGAGATGTTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGGCCC 774  
Db 245 AspGluPheSerLysAsnPheLeuProLeuGlySerGlyAsnThrAlaIleIleSerAsn 264  
QY 775 GACGCGCGCGCTGCGCGAGCGCTCCCGAGACCGAAGAGGAGCTGTGTCGCCGAT 834  
Db 265 ThrGlyGluIleLeuAlaSer---IleProGlnAspAlaGluGlyIleAlaValAlaGlu 283  
QY 835 ATGACCTCGGCATGATCGCGTTGGCCAAGCGCGCGCATCCGCGCGGCACTATTCA 894  
Db 284 IleAspLeuAsnGlnIleIleTyrGlyLysTrpLeuLeuAspProAlaGlyHisTyrSer 303  
QY 895 CCGCCGACGTAAACGCGGCTGCTGCTGATCGACGT-----CCGCGCCAAACGCGTC 945  
Db 304 ThrProGlyPheLeuSerLeuThrPheAspGlnSerGluHisValProValLysLysIle 323  
QY 946 -----GTCACGCTTGATGCCGATTCGAACCGCAAAACGAGACAAGGCGAC--- 993  
Db 324 GlyGluGlnThrAsnHisPheIleSerTyrGlyAspLeuHisGluAspLysMetAspMet 343  
QY 994 -----GCGCCGCGCGTGGCGGTG 1011  
Db 344 LeuThrIleProProArgArgVal 351

## RESULT 6

US-08-465-615-5

Sequence 5, Application US/08465615

Patent No. 5635391

## GENERAL INFORMATION:

APPLICANT: PETRE, Dominique

APPLICANT: CERBELEAUD, Edith

APPLICANT: LEVY-SCHILL, Sophie

APPLICANT: CROUZET, Joel

TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE

TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,

TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM

TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO

TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/465,615

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/194,588

FILING DATE: 10-FEB-1994

APPLICATION NUMBER: FR 9209882

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Crane-Feury, Sharon E

REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 003025-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
; LENGTH: 354 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-465-615-5

Alignment Scores:

Pred. No.:	4.76e-43	Length:	354
Score:	592.50	Matches:	137
Percent Similarity:	54.0%	Conservative:	51
Best Local Similarity:	39.4%	Mismatches:	133
Query Match:	31.5%	Indels:	27
DB:	1	Gaps:	7

US-09-751-299-1 (1-1041) x US-08-465-615-5 (1-354)

```
QY 31 GCGGCGGTGAGCGCGCGCGGTGTTCTCGATCTCGACCGCAGACTCGAGAAAGCATC 90
    |||||
Db 10 AlaAlaValGlnAlaAlaProValPheMetAsnLeuGluAlaThrValAspLysThrCys 29

QY 91 GGCCTGATCGAGCAGCGCGCAAGCAGAGCGTGCCTGATCGCATTCAGAGACTTGG 150
    |||||
Db 30 LysLeuIleIaGluAlaAlaSerMetGlyAlaLysValIleGlyPheProGluAlaPhe 49

QY 151 ATCCCGGCTATCCCTTTTGATATGGCTGGC-----GCGCCGCTTGG 195
    |||||
Db 50 IleProGlyTyrProTyrTrpIleTrpThrSerAsnMetAspPheThrGlyMetMetTrp 69

QY 196 GGCATGCGCTTCGTCCAGCGCTATTTCGAGAATTCCGCTCGTGCAGCGGAGCAGCAGTGG 255
    :
Db 70 AlaVal-----LeuPheLysAsnAlaIleGluIleProSerLysGluVal 84

QY 256 CAGGCGCTGCGCGATGCGCGCGCGCGCATGCGATGCGTGTGCGCGCTATAGCGAG 315
    |||
Db 85 GlnGlnIleSerAspAlaAlaLysLysAsnGlyValTyrValCysValSerValSerGlu 104

QY 316 CGCGCGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGCGCGCGCATGCGATCTGATC 375
    :
Db 105 LysAspAsnAlaSerLeuTyrLeuThrGlnLeuTyrPheAspProAsnGlyAsnLeuIle 124

QY 376 GCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCAGCACCCTGTTGCGGAGGAGAC 435
    |||||
Db 125 GlyLysHisArgLysPheLysProThrSerSerGluArgAlaValTyrGlyAspGlyAsp 144

QY 436 GGCAGCCATCTCGCGGTGCAGCATCCGCATCGGGCGCCTCGGCGCCTCTGTGTGTGG 495
    |||||
Db 145 GlySerMetAlaProValPheLysThrGlnTyrGlyAsnLeuGlyGlyLeuGlnCysTrp 164

QY 496 GAGCACAATCCAGCATTTGTGAAATAGCCCATGTACGCCCGCAGCAAGAGTCCACGTC 555
    |||||
Db 165 GluHisAlaLeuProLeuAsnIleAlaIleMetGlySerLeuAsnGlnGlnValHisVal 184

QY 556 GCGTCGTGCGCAGCTTC-----AGCCTCTATCGCGGCGCATGGCC 594
    |||||
Db 185 AlaSerTrpProAlaPheValProLysGlyAlaValSerSerArgValSerSerVal 204

QY 595 TATGCGCTCGGACCGGAGGTCAATACCGCGCAGCAGCATCTACCGCGTTCGAGGCGCGC 654
    |||
Db 205 CysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyrAlaIleSerAsnGln 224

QY 655 TGCTAGTGTGCGCTCGTGCAGCAGCTTTCGCGGAGATGATCAAGTATTGTGTGAT 714
    |||||
Db 225 ValTyrValIleMetSerThrAsnLeuValGlyGlnAspMetIleAspMetIleGlyLys 244

QY 715 ACGCCCGACAAGAGATGTTCTCAAGCGCGCGGCGGCTTTTGCATGATTTTCGGGCCC 774
    :
Db 245 AspGlnPheSerLysAsnPheLeuProLeuGlySerGlyAsnThrAlaIleIleSerAsn 264

QY 775 GACGCGCGCGCGCTGCGCGAGCGCGTCCGAGACCGGAAGAGGACTGCTGTGCGCCGAT 834
    |||
Db 265 ThrGlyGluIleLeuAlaSer---IleProGlnAspAlaGlnGlyIleAlaValAlaGlu 283
```

```
QY 835 ATGCACCTCGGCATGATCCGCTTGCCCAAGCGCGCGCGCATCCGGCGGCGCACTATTCA 894
    |||||
Db 284 IleAspLeuAsnGlnIleIleTyrGlyLysTrpLeuLeuAspProAlaGlyHisTyrSer 303

QY 895 CCGCCCCGACGTACCGCGCTGCTGCTGATCGACGT-----CCGGCCCCACCGCTC 945
    |||
Db 304 ThrProGlyPheLeuSerLeuThrPheAspGlnSerGlnHisValProValLysIle 323

QY 946 -----GTCACGCTTGATGCCGATTCGAACCGCAAAACGAGACAGGCGGAC--- 993
    :
Db 324 GlyGluGlnThrAsnHisPheIleSerTyrGluAspLeuHisGlyAspLysMetAspMet 343

QY 994 -----GCGCCCGCGCTGCGCGTG 1011
    |||
Db 344 LeuThrIleProProArgArgVal 351
```

RESULT 7

US-09-252-991A-26099  
; Sequence 26099, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074, 788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26099  
; LENGTH: 863  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26099

Alignment Scores:  
Pred. No.: 1.18e-13 Length: 863  
Score: 255.50 Matches: 139  
Percent Similarity: 34.1% Conservative: 9  
Best Local Similarity: 32.0% Mismatches: 160  
Query Match: 13.6% Indels: 126  
DB: 2 Gaps: 20

US-09-751-299-1 (1-1041) x US-09-252-991A-26099 (1-863)

```
QY 27 CCGCGCGCGGTGAGCGCGCGCC--GGTGTCTCGATCTCGACCGCAGTCCAGAA 83
    |||||
Db 234 ArgGlyGlyGlyGluGlyArgGlyPheGlyGlnLeuArgGlyArgArgHisArgArgVal 253

QY 84 ACGATCGG-----CCTGATCGAGCAGCGCGCAAGCAGAGAGT 122
    |||
Db 254 SerSerArgArgGlyTyrProLeuLeuLeuProGlyAspGlyTyrProProAlaGlyArg 273

QY 123 GCGCCTGATCGCATTCAGAGACTTGATTCGCGGCTATCCCTTTGGATATGCTGGG 182
    |||
Db 274 ThrPro--GlyHis-----ArgThrGlyHisArgProGlyProGlyGlyLeuAlaAsp 291

QY 183 CCGCGCGGCTTGGGCGATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGTGCGCG 242
    |||||
Db 291 roArgArgArgGlyPro-----SerA 298

QY 243 CAGCAAGCAGTGGCAGGCGCTGCGGA-----TGC-----GGCCGCGCGCCA 284
    |||||
Db 298 IaAla-AlaIaIaGlyThrGlyGlyThrAlaArgAlaCysHisArgGlyProProLeu 317

QY 285 GCGCATGATGTGTGCGCGGCTATAGCGAGCGCGCGGCGCAGCCTGTATATGGGCA 344
    |||
Db 318 ArgArgGlyArgArgSerArgLeuProAlaIaAspArgHisGlyAlaIaLeuGlyPro 337

QY 345 GCGCAT----- 350
```





```
QY 325 -----CGCCCGCGCTGCTATAGCCGCCACGACATGATCCCGTGGCGCGG 275
Db 303 LeuArgGlyLeuProAlaAlaArgArgSerArgProArgHis-----ArgArgProGly 320
QY 274 CCGCATCCCGCAGGCGCTGCCACTGCTTGCTGCCGCCACGAGCGAATTCTCGAAATAGC 215
Db 321 SerValProPro-----ValArgThrSerArgAlaThrProArgArgTrpHisAla 337
QY 214 GCTGAGCAGAGCGCATGCCCCCAAGCGCGGCCCAACCATATCCAAAGGATAGCCGG 155
Db 338 ArgTrpGlnAlaAlaIleArgArgProGly----- 347
QY 154 GAATCCAAGTCTCTGGGAATGCGATCAGCGCGCAGCTCTGCTTGCGCCGCTCGATCA 95
Db 348 -----SerValLeuSerSerAlaThrProAlaIleProGlyValArgGly 362
QY 94 ---GGCCGATCGCTTCTCGA-----CTGTGCGGTGAGATCGAGGAACA 53
Db 363 TyrProArgGlyIuTyraAlaArgAlaProArgProAlaLeuProGlyArgProAlaAlaArg 382
QY 52 CCGGCGCGCGCTGCAACCGCGCGCGCGCAT----- 23
Db 383 ProAlaArgProArgProCysArgArgProAlaAspArgSerProGlyGlnGlyArgPro 402
QY 22 -----ACTTCGTATGGGCTCCG 5
Db 403 ProGlySerSerAlaThrArgArgTrpArgPro 413
```

## RESULT 9

```
US-09-252-991A-31734
; Sequence 31734, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31734
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31734
```

## Alignment Scores:

```
Pred. No.: 7.98e-13
Score: 244.50
Percent Similarity: 37.7%
Best Local Similarity: 30.9%
Query Match: 13.0%
DB: 2
Gaps: 23
```

US-09-751-299-1 (1-1041) x US-09-252-991A-31734 (1-419)

```
QY 24 TCGCGCGCGCGC-----GGTGAGCGCGCGCGCGGTGTTCTCGA 62
Db 6 AlaArgArgGlyGlyArgArgAspAlaProHisGlyAlaGlyHisArgAlaArgArgPro 25
QY 63 TCTGCAACGCAAGTCGAGAAAGCGATCGCGCTGATCGAGCAGCGGCCAAGCAGACGT 122
Db 26 AlaArgProHis---ArgProAlaArgAlaArgAlaAspArgProGlyHisArgSerAlaArg 44
QY 123 GCGCGCTGATCGCATTCACAGAGACTTGAGATCCCGGCTATCCCTTTTGATATGCTGGG 182
Db 45 ArgProAlaAlaArgArgArg-----SerGly 53
```

```
QY 183 CGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTCGAGAAATTCGCTCGCGCGG 242
Db 54 TrpSerGlyAlaArgLeuAlaArgArgGlyCysAlaArgArg-----CysArgProAsn 71
QY 243 CAGCAAGCATGTGCGAGCGCCCTGGCGGATGCGGCCCGCGCCACGGCATGATGTCGTGGC 302
Db 72 ArgArgHisValProGlyPheProGlyAlaGlyTyraArgProArgArgThrCysSerAla 91
QY 303 CGGCTATAGCAGCGCGC-----GGCGCGCAGCTCTATATGG----- 341
Db 92 ProArgAsnProProArgArgArgTrpProGlyGlyArgArgProThrSerGlyAsnAla 111
QY 342 ---CCAGGCGATCTTTCGCGCCGATGGCGATCTGATCGCGCGCGCCGCAAGCTCAAGCC 398
Db 112 ArgProGlyProValArgAlaArgTrpArgGluProGlyCysSerPro----- 127
QY 399 TACCCATGCGGAGCGCACCGGTGTCGCGAGGAGACGCGCACCATCTCGCGGTGACGA 458
Db 128 -----GlyGlyHisArgGlyArgArgHisArgTrp-----SerArgAspArgArg 142
QY 459 TACCGCATCGGGCGCGCTCGCGCGCTCTGTGCTGGAGACATCC-----A 506
Db 143 His--ProArgGlyArgSerValArgGlnArgArgCysSerAspSerValGlyAspAsp 162
QY 507 GCCATTGCGAAATACGCCCATGACCGCCGCGACGAAACAGGTCCACGTCCGCTGTGGCC 566
Db 162 rgArgSerArgSerLeuProProAspGlyProAlaGlyArgSerGlyThrAlaProGlyA 182
QY 567 GAGCTTCAACCTCTATCGCGGCATGACCT----- 595
Db 182 rgArgSer-----ArgTrpSerGlyArgProAlaCysArgLeuArgArgAlaG 198
QY 596 -----ATGCGCTCGGACCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGC 641
Db 198 IyArgThrIleArgArgArgArgArgArgGlySerProAlaProAlaAspArgArgAlaProA 218
QY 642 GGTTCAGGCGCGCTGCTACGTGCTGGCGCTGTGCGCGCAGCGTTTCGCCGAGATGATCA 701
Db 218 rGlyS-----ThrAlaTrpArgAlaProArgHis----- 227
QY 702 GGTATTGTGATACGCGCCGACAGAGATGTTCTCAAGCGCGCGGCTTTTGGCAT 761
Db 228 -----Arg-ArgTyrglySerAlaAlaGlnGlyArgArg-----IleHis 240
QY 762 GATTTT---CGGCGCGCAGCGCGCGCTGCGCGAGCC---GCTCCCGAGACCGAAGA 815
Db 241 SerSerGlyArgCysArgArgProIleProGlyArgSerValAlaProAlaGlyArgSer 260
QY 816 GGGACT-----GCTGTGCGCGCATATGACCT----- 842
Db 261 GlyAlaGlySerValValProAlaGlyArgCysTrpArgProAsnValLeuArgCysArg 280
QY 843 -----CGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGC 881
Db 281 SerArgAsnAlaArgProGlyArgAsnAsnArgSerThrArgProThrGlyArgSerGly 300
QY 882 GGGCCACTATTACGCGCGCGCAGTAACGCG-----GCTGCTGCTGGA---TCG 926
Db 301 SerThrProGlyAlaAlaArgArgArgAlaGlyArgArgPheAlaSerAlaGlyArgAla 320
QY 927 ACGTCCGCGCCCAACGCGTGTCAAGCTTGATGCGCG----- 962
Db 321 GlyAlaGlyLeuArgArgArgSerGlyCysCysArgProGlyArgAlaProSerSerHis 340
QY 963 -----ATTGCAACCGCAAAAGAGACAAGGCGGCGCGCGC----- 1001
Db 341 CysProValArgArgAlaArgArgSerSerGlyAlaAlaSerArgProProAlaArgPro 360
QY 1002 GCTGCGCGTGTGTGCGGAAAGCGCGCGCGCGCA 1037
Db 361 AlaAlaArg-----SerLeuArgArgArgAla 369
```

## RESULT 10

```
US-09-252-991A-31715
; Sequence 31715, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31715
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31715

Alignment Scores:
Pred. No.: 1.48e-12 Length: 724
Score: 242.50 Matches: 137
Percent Similarity: 31.2% Conservative: 22
Best Local Similarity: 26.9% Mismatches: 144
Query Match: 12.9% Indels: 207
DB: 2 Gaps: 26

US-09-751-299-1 (1-1041) x US-09-252-991A-31715 (1-724)

QY      1012 CCACGCCGAGCGGGCCGCTGCCT-----TGTCTCGTTTGCGGTTCGAATGCCG 959
        |||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       70 ProGlnValGlyHisAlaValArgProGlnHisLeuProArgLeuGlnArgLeuGln 89

QY      958 CATCAAGCGTGACGACGCGTTGGCCGAGCGTGCATCCAGCAGACCGCGTTACGTCGG 899
        ||||| |
Db       90 HisGlnPro-----ArgArg 94

QY      898 GCCGTGAATAGTGCCCCCGCATCGCCGCG----- 866
        |||::: ||||| ||||| |||||
Db       95 AlaLeuArgArgAspProProHisArgGluProLeuGlnArgLeuGlnProLeuLeu 114

QY      865 CTTGGCCAACGCCGATCATGCCAGAGTCGATATCGGCCAGCACAGATCCCTCTCGGTCT 806
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       115 ProLeuArgProArgSerLeuArgLeuArgLeuArgLeuArgGlnProArgArgPro--SeraAla 133

QY      805 CCGGAGCGGCT-----CGGCCAGGG 785
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       134 ProAlaAlaAlaGlyAlaGluGlnAlaGluAlaGluProGlyArgArgProArgProGly 153

QY      784 CGCGGC----- 779
        |||||
Db       154 ArgArgProAlaGluGlyAlaGlnGlyGlyArgHisArgPheAlaAlaArgGlnProGly 173

QY      778 -----CGTCGGGCCCAAATCATGG-----CAAAACCGCCC 746
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       174 LysGlnLeuArgProAlaArgAlaGlyTyrArgArgGlnLeuLeuGlnArgHisGlnArg 193

QY      745 CGGCCTTAGGAACATCTCTTTCGGGGCGTATCCACCATACTTGATCATCTCCGGCG 686
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       194 Arg-----ArgAlaArgProProAlaPro-----AspProAla 204

QY      685 AAACGGTCGCGCAGCAGCCAGCAGTAGCAGCCGCCCTGCAGCGGT-----AGATCT 632
        ::: ||||| ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       205 GlyHisAlaGlyArgProAlaAlaGlyAlaGlnHisProArgHisArgGlyProArgArg 224

QY      631 GCCTTCGGCGGTAT-----TGACCTCCGGTCCGAGCCGATAGGCCA----- 590
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db       225 GlyLeuArgProArgArgGlyProAspProAspArgArgProHisArgProGlyProAla 244

QY      589 -----TGCCGCGATAGAGGCTGAAGCTCGGCC 563
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db      245 ProvalGlyGlyGlnGlyArgArgAspGlyCysArgHisGlyGlyProAlaValAla      264
Qy      562 ACGACGCGA-----                    554
Db      265 ArgArgArgGlyGlyGlnHisArgProAlaCysAlaGluProAlaValHisArgGlnArg      284
Qy      553 ---CGTGAACCTGTTGTCGCGCGCGGTACATGGCGTATTTCGACAAATGGCTGATGTGCT      497
Db      285 GlnArgAspProProArgArgArgGlyArgGlyAsnArgSerArgArgAlaGlyArgSer      304
Qy      496 CCCAGCAACAGAGCGCGCGCGCCCGATGGCGGTATCGTGACCG-----          449
Db      305 ArgSerProArgLeuArgArgGlyProArgAspArgProLysArgProValArgGlyArg      324
Qy      448 -----CGAGATGGCTGCCGCTCCCTCCCGCCGAACACGCGTGCGCTCCGATGGG      401
Db      325 ProArgProAlaGlyProGlyLeuArg--ProAlaHisArgArgCysProAlaGlyGly      343
Qy      400 TAGCCTTAGCT-----TGGCGCGCGCGCGGATCAGATCGCCATCGG      359
Db      344 GlnAlaSerAlaGlySerValArgGlnLeuAlaGlyGlnGlnGlyLeaAspArgSerArg      363
Qy      358 GCGCGAAGATCGCCTGGCCCA-----                    338
Db      364 GlyGlnHisArgGlnGlyProGlnAlaAlaArgGlnGluArgPheHisGlnProGlyGly      383
Qy      337 -----                    320
Db      384 ArgArgSerGlnGlnProArgProGlyAlaLeuArgArgArgLeuArgGlySerArgAla      403
Qy      319 CGCGCT-----CGCTATAGCCGCGCACGA-----          296
Db      404 ArgAlaThrAspLeuArgThrGlyArgArgArgGlyGlyAlaGlyGluArgSerLeuSer      423
Qy      295 -----CATGCATGCCGT-----GCGCGCGGG-----          275
Db      424 ProHisArgHisArgProCysArgArgGlyProGlyGlyGlnGlyGlyAspArgArg      443
Qy      274 -----CCGCATCCGCCAGGCGCTGCCACTGCTTGCTGC      242
Db      444 ArgSerProAlaAspArgHisHisArgGlnGlyAlaProGlyAlaAlaGlyGlnLeu      463
Qy      241 CGCGCAGAGCGAATTCGAAATAGCGCTGACGAAGCGCATGCCCAAGCGCGG---      185
Db      464 ArgArgArg-----ArgArgHisProGlyGlnProGlyArgProArgProAlaLeu      480
Qy      184 -----CGCCACGCATATCCAAAAGGATAGCCGGGAATCCAACTCTTGGA      137
Db      481 LeuProGlyValArgProAspLeuLeuArgArg-----GlnGlnHisGlyProArgGly      498
Qy      136 ATGCATCAGCGCGCACGCTCTGCTTGCGCGCTGCTGCATCAGGCGGATCGCTTCT---      80
Db      499 LeuAlaLeuAlaAlaCysAspProGlnHisProAlaArgGlnAlaArgArgLeuAspPro      518
Qy      79 -----CGACTGTGCGGTGAGATCGAGATCGAAGAACCGGCGCGCTGCACCG      35
Db      519 ValGlyProArgHisArgSerArgGlyArgGlyLysAlaAspProGlyArg-----His      536
Qy      34 CCGCGCGCGGATACTTCGTTCATGGGCTCCG      5
Db      537 ProArgArgArgThrGlyArgLeuValPro      546

RESULT 11
US-09-252-991A-23569
; Patent No. 6551795
; Sequence 23569, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18

```



US-09-751-299-1 (1-1041) x US-09-252-991A-25394 (1-369)

```
QY 1039 ACTGCGCGCGCGCGCTTCCG-----CCACCACGCGCAGCG----- 1001
Db 75 ThrLeuArgArgArgTrpProGlySerArgProProArgSerSerThrGlyProGly 94
QY 1000 -----CGGCGCGTCCCTTGCTGCTTTGCGGTGCAATGCGCATCAA 953
Db 95 ArgArgProProArgArgArgProProArgArgSerAlaAlaAlaSerProArgHisArg 114
QY 952 GCGTGACGACGCGTTGGCCGAC---GTGATCCAGCAGCAGCGCGGTTACGTCGGCC 896
Db 115 ThrSerAlaProArgGlyProGlyArgArgThrProAlaSerSerSerGluArgProGly 134
QY 895 GTGAATAGTGGCCCGCCGATCGGCCCGCT---TGCCACACCGCATCATCGCAGGT 839
Db 135 ThrAlaProAlaProProArgArgThrAlaProGlyTrpProGlyProSerAlaAlaGly 154
QY 838 CGATATCCGCGCAGCAGCAGTCCCTCTTCGCTCCGCGGAGCGGCTCGG----- 791
Db 155 AlaGlyArgProAlaArgArgProAlaArgPheProProProArgArgCysArgIlePro 174
QY 790 -----CCAGGCGCGCGCGCGTCCGCGCGCAAAATCATGCAAAACCGCGCGG----- 743
Db 175 AlaGlyProGlyProProArgArgProGlyArgProTrpArgGlyThrAlaArgTyrArg 194
QY 742 -----CCTTGAGAACATCTCCTTGTGCGGCGTAATCCACCAATACCT 701
Db 195 CysArgGlyCysCysAlaPro-----AlaProArgArgArgSerProAlaIlePro 211
QY 700 TGATCATCTCCGCGGAAACGGTCCGCGCAGCAGCAGCATAGCAGCCGCTCGACCG 641
Db 212 AspAlaArgProAlaCysAlaAlaProArgAlaGlyAlaGlyAlaAlaArgProAlaGly 231
QY 640 CGTAGATCTGGCTTCCGCGCGTATTGACCTCCGCTCCGAGCGCATAGCCATGCGCGAT 581
Db 232 GlnglySer-----ProAlaArgCysAlaAlaProSerArgPro 244
QY 580 AGAGGCTGAAGCTCGGCCAGCAGCAGCGAGTGAAGCTGTTGTCGCGCGGTACATGGCGT 521
Db 245 AlaAlaAlaArgSerAlaThrThrGlyArgArgPro-----SerAlaSerArg 260
QY 520 ATTTGACAATGGCTGATGTGCTCCAGCAACAGAGCGCGCGGCGCCGATGCGCG 461
Db 261 AlaGlyArgProAlaArgCys---ProAlaGluArg-ArgAlaThrAlaPro----- 276
QY 460 TATCGTGACCGCGAGATGGCTGCGCTCCCTCCGCGAACAAGGCTCCGATGCG 401
Db 277 -AlaAlaHisArg-AlaTrp---ProProProAlaPro-AlaProAlaAlaProAla--- 293
QY 400 TAGGCTTGAGCTTGGCGCGCGCGCATGATCGCCATCGGGGCGGAGATGCGCTGGC 341
Db 294 -----AlaProGlyAlaAlaArgSer-----ArgProAlaGly 304
QY 340 CCATATAGAGGCTGCGCGCGCGCTCGCTATAGCCGCGCAGCATGCGATGCGTGC 281
Db 305 ProArgArgGlyArgArgArgArg-----ArgCysArgSerAlaProAlaGly 320
QY 280 GCGGCGCGCATCCGCCAGGCGCTGCACTGTGCTGCGCGCAGCAGCGAATTCTCGA 221
Db 321 SerGlyGlyAsnProArgGlySerArgGlyGlyCysCysBarArgArg----- 336
QY 220 AATAGCGGTGACGAGCGCATGCGCCAGCGCGCGCGCGCCAGCCATATCCAAAGGAT 161
Db 337 -----ArgSerAlaAlaProAlaAlaAlaProGlyAlaAlaAla----- 349
QY 160 AGCCGGAAATCCAAGTCTTGGAATGCGATCAGGCGCAGCTCTGCTTGCGCGCTGCT 101
Db 350 -----SerProGlySerGlySerProProAlaAlaArgProValAlaAsnProArg 366
QY 100 CGA 98
|||
```

```
Db 367 Arg 367
RESULT 13
US-09-252-991A-22486
; Sequence 22486, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22486
; LENGTH: 344
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22486

Alignment Scores:
Pred. No.: 2.45e-12 Length: 344
Score: 238.50 Matches: 114
Percent Similarity: 36.8% Conservative: 20
Best Local Similarity: 31.3% Mismatches: 120
Query Match: 12.7% Indels: 110
DB: 2 Gaps: 21

US-09-751-299-1 (1-1041) x US-09-252-991A-22486 (1-344)
QY 1034 GCGCGCGCGCGCTTCCGCCACCAAGCGCAGCGCGCGCGCTTGTCTCTGTTT 975
Db 1 AlaGlySerProValArgArgArgIleArgSerAspAlaProSerPro----- 16
QY 974 TGGCGTTGAATGCGGCATCAAGCGTGACGAGCGGTTGGCCGCGAAGTCGATCAGCAGC 915
Db 17 CysArgArgArgArgSerAlaIleAlaArgThrArgTrp-----ArgArgSerPro 34
QY 914 AGCCGCTTAAGTCCGCGCGCTGAATAAGTGCCCGCGCGGATGCGCGCCGCTTGCCCAAC 855
Db 35 SerProGlySerThrGlyArgGly---Trp-ArgProGlyArg----- 47
QY 854 GCGATCATGCCAGTGCATATCGCGCAGCAGCAGTCCCTCTGCTCCGCGAGCGGC 795
Db 48 -----ArgLeuArgAlaProAla-----ArgSerProSerAlaAl 59
QY 794 TCGGCCAGGCGCGCGCGCTCGGCGCCGAAATCATGGCAAAACCGC----- 749
Db 59 aaArgAlaGlyArgTrpArgArgAspArgSerGlyCysArgThrArgArgTrpArgSerIle 79
QY 748 -----CGCCGCGCTTGAGGAACATCTCC 726
Db 79 uArgGlySerAlaArgGlyValAlaAlaArgArgGlyCysArgArgArgCysGlyThrSerPr 99
QY 725 TTGTGCGCGCTATCCACCATACTTGATCATCTCCGCGCAACGGTCCGCGCAGCAGCC 666
Db 99 oglyArgAlaAlaArgSer-----SerGlyProCysSerArgThrGlySerSerAl 116
QY 665 AGCAGCT---AGCAGCGCGCTCGACCGCGTAGATCTGCGCGGTATTGACCTCC 609
Db 116 aaArgArgProSerAlaArgProGlyProAlaArgArgGly----- 129
QY 608 GGTCCGAGCGCATAGGCGATCGCGCGCATAGAGGCTGAAGCTCGGCCACGACGACGTGG 549
Db 130 -----SerAlaArgArgSerCysArgGly----- 137
QY 548 ACCTGTTCTGTCGCGCGGTACATGGCGTATTTCGACAATGCGCTGATGTGCTCCAGCAA 489
Db 138 -----ArgArgGlyArgTrpArgArgSerSerTrpArgTyrCysAlaArg----- 152
```



QY 488 CAGAGCGCGCGAGCGCCGATGCGGTATCGTCACCGGAGATGCGTCCGCTCTCCC 429  
Db 153 -----ArgTrpSerThrAlaArgThrGlyCysSerValPr 164  
QY 428 T-----CGCCGAACACGCGTGCCTCGCATGGTAGGC 396  
Db 164 oSerGlyAlaLeuArgProTrpSerGlyArgAspArgArgCysThrGlyAlaGlySerAl 184  
QY 395 TTGAGCTTGCGCGCGCGCGCATCAGATCGCCATCGGGC-----CG 354  
Db 184 aAlaAlaValArgGlyArgArg---AspGlyTrpArgGlyAlaGlyAsnArgAspTrpAr 203  
QY 353 AAGATCGCTGCGCCCATATAGAGCTGCGCGCGCTCGCTATAGCCGCGCAGACA 294  
Db 203 gGluCysProGlyPro-----GlyAlaProProArgThr-----ProLeuAr 217  
QY 293 TGCATGCGCT-----GGCGCGCGCGCATCCGCGAGGCGCTGCCACTGCTTGTG 243  
Db 217 gAlaAlaArgArgValAlaGlySerArgAspTrpProGluCysProGlyArgAla----- 235  
QY 242 CCGCGCAGAGGAATTCTCGAATAGCGGTGACGAGCGCATGCCCAAGCGCGCG 183  
Db 236 ----ValArgArgArgThrArgArgSerAlaAspArgArgAlaAlaGlyCysArgGlyPh 254  
QY 182 CCCAGCCATATCCAAAGGATAGCCCGGAATCCAACTCTTGGAATGCGATCAGGCGC 123  
Db 254 eProGlyCysArgGlyArgAlaAlaArgGlu---LysThrArgGlyAlaArgArgAlaAl 273  
QY 122 AGCTCTGCTTGG-----CCGCGTGTCT-----CGATCAGCGCGATCGCTTTC 81  
Db 273 aSerGlyCysTrpThrTrpArgSerProProAlaValAlaGlyArgThrArgArgGluValGl 293  
QY 80 TCGACTGTGC 71  
Db 293 ypheValCys 296

RESULT 14  
US-09-252-991A-24567  
; Sequence 24567, Application us/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074, 788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094, 190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24567  
; LENGTH: 681  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24567

Alignment Scores:  
Pred. No.: 3.21e-12 Length: 681  
Score: 238.50 Matches: 126  
Percent Similarity: 33.9% Conservative: 19  
Best Local Similarity: 29.4% Mismatches: 152  
Query Match: 12.7% Indels: 131  
DB: 2 Gaps: 20

US-09-751-299-1 (1-1041) x US-09-252-991A-24567 (1-681)

QY 1036 GCGCGCGCGCGCGCTTTCGCCACGCGCAGCGCGCGCTGCG----- 989  
Db 79 SerLysArgSerAlaProProAlaProArgAlaSerTrpSerThrArgThrAlaArgGly 98

QY 988 ---CCTTGT-----CCTCGTTTTCGGTTCGAATGCGCATCAAGCG----- 950  
Db 99 SerProCysAlaTrpMetProLysAlaAlaCysArgArgArgProAlaAlaAlaMetArg 118  
QY 949 -----TGACGACCGCTTGCGCGACGTC 926  
Db 119 SerCysArgTrpSerSerGlyAsnGlyArgTrpArgTrpArgArgAlaValArgPro 138  
QY 925 GATCCAGCAGCAGCGCGCTTACGTGCGCGCGTGAATAGTGGCCCGCGGATCGCGCGCG 866  
Db 139 TrpProAlaSerProArgValAlaArgValAlaGlyValTrpProProSerLeuCysPro 158  
QY 865 CTTGGCCCAACGCGATCATGCGAGGTGATATCGGCGCAGCAGCATCCCTCTT----- 812  
Db 159 AlaProProArgArgArgProGlyArgGlnArgArgProArgGlyProAlaAlaGln 178  
QY 811 -----CGGTCTCCGGGA-----GCGGCTCGGCGCAGCGCGCG 779  
Db 179 ArgArgAlaAlaTrpArgArgArgAlaGlyAspGlnProAlaAlaArgProGlyArgGly 198  
QY 778 CGTGG-----GCC 770  
Db 199 GlnArgThrCysLeuGlnProLeuLeuAlaValGlnProAlaValLeuGlnCysProAla 218  
QY 769 CGAAATCATGCGCAAAACCGCGCGCGCTTGAGGAACATCT----- 728  
Db 219 ArgArgAlaGlyAspHisSerArgArgGlySerGlyGlyThrGlyAspProSerArgArg 238  
QY 727 CCTGTGCGGCGGTATCCACCAATACCTTGATCATCTCGCGCGAAACG----- 680  
Db 239 ProGlyArgGlyAspGlyPro---ProGlyGluProGlyAlaAspArgLeuAspSerGly 257  
QY 679 -----TCGCGCAGCAGCGCAGCAGTAGACGCC----- 650  
Db 258 GlyGlyProAlaHisAlaProAlaAlaProThrProProArgLeuHisArgAlaArgGln 277  
QY 649 -----CCTGACCGCGGTAGATCTGCTTGGCGCGGTATTGACCTCCGCTCCGACG 599  
Db 278 SerLeuAlaProArgProArgArgIle-----ProProArgSer 290  
QY 598 CATAG-----CCATCGCGGATAGAGCTGAAGCTCGGCGCACGACG 557  
Db 291 ArgArgGlyProAlaAlaProLeuProLeuArgAsp-----Pro 303  
QY 556 CGAGGTGACCTGTTGTCGCGCGCGGTACATGCGGTATTGACAATGCTGTGATGCT 497  
Db 304 AlaGlyAlaProGlyArgArgPro-----ArgLeuGluThrLeuAlaGlyAlaAla 320  
QY 496 CCCAGCAACAGAGCGCGCGAGCGCGCGATGCGGTATCGTCACCGCAGATGCGTGC 437  
Db 321 AlaProProArgArgAlaGlyGlyArg---ValLeuArgArgSerArgArgGlySer 339  
QY 436 CGTCTCCCTCGCGCAACAGGTGCGCTCCGATGGGTAGGCTTGAGCTTGC----- 386  
Db 340 GlyLeuProCysLeuArgProValAlaAspProThrLeuProAlaThrCysProAlaPro 359  
QY 385 GCGCGCGCGCGATCAGATGCGCATCGGCGCGGCGGAGATGCGCTGAGCGCTGC 326  
Db 360 GlyAlaArgGlyArgAspGlyHisArgSerGlyArgArgProGly-----GlyGly 376  
QY 325 CGCGCGCGCGCTCGCTAT-----AGCGGCGCAGCATGCA 290  
Db 377 ArgArgArgArgArgGlnProGlyLeuValAlaAlaGlyArgAlaAlaArgArgSerGlu 396  
QY 289 TCGCGTGGCGCGCGCGCATCGCGCAGGCGCTGCTTGTGCGCGCGCAGCAGCG 230  
Db 397 CysArgArgAlaAlaGlyHisProGlnProValArgThrGlyLeuGlyArgPheArgVal 416  
QY 229 AATTCTGAAATAGCGCTGACGAGCGCATGCCCAAGCGCGCGCGCGCATATCC 170  
Db 417 GlnSerArgArg-----ProAlaProProArgLeuProGlyIle 429  
QY 169 AAAAGGATAGCGCGGAATCCAAAGTCTCTGGGAATGCGATCAGGCGCAGCTCTGCTTG 110

```
Db      430 ProArgSPAlaIaGlyGlnProGlyLeuAlaArgAlaThrHisArgProCysAsp 449
QY      109 CCGCTGCTGCATCAGCGCGATCGCTTCTGACTGTGCGGTGAGATCGAGAACACCG 50
Db      450 GlyProAlaThrSer-----LeuAlaAspPro 458
QY      49 GCGCGCGCTGCACCGCGCGCGC 26
Db      459 AlaArgProAlaAlaProAlaArg 466

RESULT 15
US-09-252-991A-16655
; Sequence 16655, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16655
; LENGTH: 977
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16655
```

```
Alignment Scores:
Pred. No.:      5e-12      Length:      977
Score:          237.00     Matches:      128
Percent Similarity: 32.0%   Conservative: 19
Best Local Similarity: 27.8% Mismatches:    155
Query Match:     12.6%     Indels:      158
DB:              2        Gaps:       22
```

US-09-751-299-1 (1-1041) x US-09-252-991A-16655 (1-977)

```
QY      1009 CGCGCAGCGCGGCGGCTGCGCTTGTCTGTTTGGCGTTGGAATGCGGCATCAAGCG 950
      |||      |||      |||      |||      |||      |||
Db      280 ArgArgGlyArgGlyGlyHisProLeuArgArgGlyAlaArgArgHisArgHisProGlu 299
QY      949 TGACGACGCGTTGGCGCGGACGTCGATCCAGCAGCAGCGCGGTACGTCCGGCCGTG--- 893
      ::|||      |||      |||      |||      |||      |||
Db      300 ArgGlnArg-----ProAspGlyAlaGlyAlaAlaProAlaArgGlyIleAlaArg 315
QY      892 -----AATAGTGCGCGCGCG----- 878
      |||      |||      |||      |||      |||      |||
Db      316 ProLeuGlyGlnArgGlnArgSerGlyAlaProGlyArgSerArgThrGlnArgGlnGly 335
QY      877 -----GATCGGCGCGCGCTTGGCCAAAGCGCATGACCGCA 842
      |||      |||      |||      |||      |||      |||
Db      336 AlaThrProGlyAlaValaArgLeuAspHisHisProGlnGlnProAlaArgAlaArgArg 355
QY      841 GGTGCA----- 836
      |||      |||      |||      |||      |||      |||
Db      356 AspArgGlnHisThrGlyAspGlnProSerGlyGlyGlyHisArgArgArgGlnProHis 375
QY      835 -----TATCGCGCA-----CCAGCA 821
      ::|||      |||      |||      |||      |||      |||
Db      376 ArgArgAlaThrAspGlyGlyArgArgGlnHisArgArgAlaGlnAspAspGlnProLeu 395
QY      820 GTCCCTTTCGGTCTCCGGGAGCGGCTCGGCCAGGCGCGCGTCCGGGCCGAAATCA 761
      ::|||      |||      |||      |||      |||      |||
Db      396 LeuLeuLeuArgProProGlyGlyArgArgAspGlyArgGly-----GlyLeu 411
QY      760 TGGCAAAACCGCGCGCGCTTGAGAACATCTCTGTGCGGCGTATCCA----- 710
      |||      |||      |||      |||      |||      |||
```

```
Db      412 HisProGlyArgAlaArgProAlaArgThrSer---CysHisProValProGlyValSer 430
QY      709 -----CCAATACCTTGATCATCTCCGGCGAAACGGTCGCGCAGCAGC----- 668
      |||      |||      |||      |||      |||      |||
Db      431 AspGluProAspProGluAspPheProAlaAspArgArgArgSerArgArgLeuArg 450
QY      667 -----CCAGCAGCTAGCAGC-----CGCCCT 647
      |||      |||      |||      |||      |||      |||
Db      451 ArgGlyAspProCysArgAlaThrGlyHisSerHisArgThrGlyGlyArgArgPro 470
QY      646 CGA-----CCGCGTAGATCT 632
      |||      |||      |||      |||      |||      |||
Db      471 ArgArgHisLeuProGluArgArgLeuHisProValGluGlyAlaAspProAlaArgArg 490
QY      631 GCGTTGCGCGGTATTGACCT-----CCGTCGAGCG----- 599
      |||      |||      |||      |||      |||      |||
Db      491 GlyIleProGlnGlyProProLeuCysArgProValGlyAlaGlyHisProGlyThrGly 510
QY      598 -----CATAGCCATGCGCGGATAGAGGCTGAAGCTCGGCCAGCAGCAGCT 551
      |||      |||      |||      |||      |||      |||
Db      511 AlaGlnHisArgHisArgProHisArgGly-----MetGluGlyArgHisArgArg 527
QY      550 GACCTGTCTCGCGCGCGGTACATGCGGTATTTCGACAAATGGCTGGATGTGCTCCAGC 491
      |||      |||      |||      |||      |||      |||
Db      528 ProProHisGlnArgArgArgArgAlaAlaGluGluThrArgGlyArgCysArgProGly 547
QY      490 -----AACAGAGCGCGCGCAGCGCGCA----- 467
      |||      |||      |||      |||      |||      |||
Db      548 LeuGlyGluAspProAlaArgArgGlnLysArgGlyGlyArgThrArgArgArgGlnPro 567
QY      466 ---TGCGGTATCGTGCACCGCGAGATGCTGCGTCTCCCTCGCGCAACACGCTGCGCT 410
      |||      |||      |||      |||      |||      |||
Db      568 AlaHisArgValaArgAlaProAlaAlaGlyArgArgLeuAlaGluArgAlaThrAsp 587
QY      409 CCGCATGGGTAGGCTTGAGCTTGCAGCGCGCGCGGATCAGATCGGCATCGG-----GCG 356
      |||      |||      |||      |||      |||      |||
Db      588 Pro-----AlaAlaGlyArgGlnGlyAspLeuLeuHisArgGlyAlaGly 602
QY      355 CGAAGATCGCT-----GGCCCATATAGAGGCTGCCCGCGCGCT 314
      |||      |||      |||      |||      |||      |||
Db      603 AlaArgGlnProAlaGlnAlaProGlyGlyArgArgArgLeuHisArgProGlyAla 622
QY      313 CGGTATAGCGCGCCACGACATGCATGCGCGCGCGG-----CCGCAT 269
      |||      |||      |||      |||      |||      |||
Db      623 GlyThrArgLeuProGlnAlaArgArgArgGlyGlyGlyGlySerAlaThrAlaHis 642
QY      268 CCGCCAGGCGCTGCCACTGCTTGTGCGCGCGCAGCAGCGCAATTCTCGAAATAGCGCTGCA 209
      |||      |||      |||      |||      |||      |||
Db      643 ProAlaGlyLeuArgArgArgArgThrAspGlnAlaGlyGlyProGlyLeuAlaGlnAlaGly 662
QY      208 CGAAGCGCATGCCCAAG-----CCGGCGCGCCAGCCATATCCAAAGGATAGCGCG 155
      |||      |||      |||      |||      |||      |||
Db      663 ArgArgAlaValaProArgAlaGlnProAlaGlyProGluArgGluArgArgAlaGlyPro 682
QY      154 GAATCCAAAGTCTTGGAATGCGATGCGAGCGCAGCTCTGCTTGGCCCTGCTGCATCA 95
      ::|||      |||      |||      |||      |||      |||
Db      683 ArgArgArgArgArgGlyAlaAlaArgAspArgArgArgProGlyThrGlyGlyArgPro 702
QY      94 GAGCGATCGCTTCTGCACTGTGCGGTGAGATCGAGGAACACCGCGCGCTGCACCG 35
      ::|||      |||      |||      |||      |||      |||
Db      703 GlnAlaAlaGlnArgArgLeu-----GluProGlyLysProGlyPro 716
```

Search completed: April 27, 2006, 01:40:53  
Job time : 49.473 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:38:31 ; Search time 32.8257 Seconds  
(without alignment)  
2650.121 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1881  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 segs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090042\_18780/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_Main -QFM=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US09751299\_@CGN\_1\_1\_405\_@runat\_26042006\_090042\_18780 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1806	96.0	346	3	US-09-751-299-2
2	1806	96.0	346	4	US-10-146-772-384
3	1806	96.0	346	4	US-10-241-742-384
4	1806	96.0	346	4	US-10-440-523-384
5	1806	96.0	346	4	US-10-440-503-384
6	1806	96.0	346	4	US-10-461-925-384
7	1464	77.8	341	4	US-10-146-772-34
8	1464	77.8	341	4	US-10-241-742-34
9	1464	77.8	341	4	US-10-440-523-34
10	1464	77.8	341	4	US-10-440-503-34
11	1464	77.8	341	4	US-10-461-925-34

12	1209	64.3	348	4	US-10-146-772-16	Sequence 16, Appl
13	1209	64.3	348	4	US-10-241-742-16	Sequence 16, Appl
14	1209	64.3	348	4	US-10-440-523-16	Sequence 16, Appl
15	1209	64.3	348	4	US-10-440-503-16	Sequence 16, Appl
16	1209	64.3	348	4	US-10-461-925-16	Sequence 16, Appl
17	1165	61.9	353	4	US-10-146-772-212	Sequence 212, App
18	1165	61.9	353	4	US-10-241-742-212	Sequence 212, App
19	1165	61.9	353	4	US-10-440-523-212	Sequence 212, App
20	1165	61.9	353	4	US-10-440-503-212	Sequence 212, App
21	1165	61.9	353	4	US-10-461-925-212	Sequence 212, App
22	1155	61.4	353	4	US-10-146-772-232	Sequence 232, App
23	1155	61.4	353	4	US-10-241-742-232	Sequence 232, App
24	1155	61.4	353	4	US-10-440-523-232	Sequence 232, App
25	1155	61.4	353	4	US-10-440-503-232	Sequence 232, App
26	1155	61.4	353	4	US-10-461-925-232	Sequence 232, App
27	1148	61.0	333	4	US-10-146-772-350	Sequence 350, App
28	1148	61.0	333	4	US-10-241-742-350	Sequence 350, App
29	1148	61.0	333	4	US-10-440-523-350	Sequence 350, App
30	1148	61.0	333	4	US-10-440-503-350	Sequence 350, App
31	1148	61.0	333	4	US-10-461-925-350	Sequence 350, App
32	1132.5	60.2	358	4	US-10-146-772-170	Sequence 170, App
33	1132.5	60.2	358	4	US-10-241-742-170	Sequence 170, App
34	1132.5	60.2	358	4	US-10-440-523-170	Sequence 170, App
35	1132.5	60.2	358	4	US-10-440-503-170	Sequence 170, App
36	1132.5	60.2	358	4	US-10-461-925-170	Sequence 170, App
37	1130	60.1	354	4	US-10-146-772-102	Sequence 102, App
38	1130	60.1	354	4	US-10-241-742-102	Sequence 102, App
39	1130	60.1	354	4	US-10-440-523-102	Sequence 102, App
40	1130	60.1	354	4	US-10-440-503-102	Sequence 102, App
41	1130	60.1	354	4	US-10-461-925-102	Sequence 102, App
42	1125	59.8	345	4	US-10-146-772-334	Sequence 334, App
43	1125	59.8	345	4	US-10-241-742-334	Sequence 334, App
44	1125	59.8	345	4	US-10-440-523-334	Sequence 334, App
45	1125	59.8	345	4	US-10-440-503-334	Sequence 334, App

ALIGNMENTS

RESULT 1  
US-09-751-299-2  
; Sequence 2, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; CURRENT FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
; US-09-751-299-2  
  
Alignment Scores:  
Pred. No.: 6.43e-123  
Score: 1806.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.0%  
DB: 3  
Length: 346  
Matches: 346  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0



US-09-751-299-1 (1-1041) x US-09-751-299-2 (1-346)

```
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGCTGCAGCGCCGCGGTGTTCTC 60
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20
QY 61 GATCTGACCCGACAGTCGAGAAAGCGATCGGCTGATCCGAGCAGCGCGCCACAGCAGAC 120
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40
QY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATTTCCGGCTATCCCTTTGGATATGCTG 180
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAAATTCGCTGCGC 240
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80
QY 241 GGCAGCAAGCAGTGGCAGGCGCCCTGCGGATGCGCGCCCGCCAGCGCATGCTGCTG 300
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100
QY 301 GCGCGCTATAGCAGCGCGCGCGCGCGCTTATATGGGCCAGGCGCATTTTCGGGCCCC 360
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120
QY 361 GATGCGCATTTGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCATCGCGAGCGCACCGTG 420
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140
QY 421 TTCGGCGAGGAGACGGCAGCCCATTCGCGGCTGCACGATACCGCATCGGCGCGCTCGC 480
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160
QY 481 GCGCTCTGTTGCTGGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180
QY 541 GAACAGGTCCACGTGCGCGCTGCGCGAGCTTCAGCCTTATCGCGGCGCATGCGCTATGCG 600
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200
QY 601 CTCGGAACGAGGTCAATACCGCGCGAGCCAGATCTACGCGGTGAGGGCGCTGCTAC 660
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220
QY 661 GTGCTGCGCTGCTGCGCGACCGTTTCGCCGAGATGATCAAGTATTGGTGATACGCCC 720
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGGCGCGACGCG 780
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260
QY 781 CGCGCCTTGGCCGAGCGCTCCCGGAGACCGGAAGAGGACTGCTGTCGCCGATATGAC 840
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValAlaAspIleAsp 280
QY 841 CTCGGCATGATCGCGCTTGGCCAAGCGCGCGCGCGCATCCGCGCGGCCACTATTACGCGCC 900
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300
QY 901 GACGTAACGCGGCTGCTGCTGATCGACGTCGCGGCCCAACGCGTGTGACGCTTGATGCC 960
Db 301 AspValThrArgLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320
QY 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGCGCGCTGCGGTGCGCGAA 1020
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340
QY 1021 AGCGCGCGCGCGCGCGCAG 1038
|||||
```

```
Db 341 SerAlaAlaAlaGln 346
RESULT 2
US-10-146-772-384
; Sequence 384, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PR1
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-384
Alignment Scores:
Pred. No.: 6.43e-123 Length: 346
Score: 1806.00 Matches: 346
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
Query Match: 96.0% Indels: 0
DB: 4 Gaps: 0
US-09-751-299-1 (1-1041) x US-10-146-772-384 (1-346)
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGCTGCAGCGCCGCGGTGTTCTC 60
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20
QY 61 GATCTGACCCGACAGTCGAGAAAGCGATCGGCTGATCCGAGCAGCGCGCCACAGCAGAC 120
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40
QY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATTTCCGGCTATCCCTTTGGATATGCTG 180
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAGAAATTCGCTGCGC 240
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80
QY 241 GGCAGCAAGCAGTGGCAGGCGCCCTGCGGATGCGCGCCCGCCAGCGCATGCTGCTG 300
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100
QY 301 GCGCGCTATAGCAGCGCGCGCGCGCGCTTATATGGGCCAGGCGCATTTTCGGGCCCC 360
|||||
```

Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCATCTGATCGCCGCGCCGAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgGlyLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGGACCATCTCGCGGTGCACGATACCGGCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGGAGACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTCGCGCTGTCGGCGGAGCTTCAAGCTCTATCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGGAGCCAGATCTACGCGGTGCGAGGCGGCTGCTAC 660  
Db 201 LeuGlyProGluValAlaSerThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGGTATTGGTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGGCGCGGCGGCTTTGCCATGATTTTCGGGCGCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGGCGCTGGCGGAGCGGCTCCCGGAGACCGAAGAGGAGTGTGTCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGCGGCGGCACTATTCAAGGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAAAGCGGCTGCTGTCGATCGACGTCCGCGGCAACGCGTCTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGACACAGGCGCGACGCGCGCGCTGCGCGTGTCGGCGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 3  
US-10-241-742-384  
; Sequence 384, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitric Oxide Synthase  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-384

Alignment Scores:  
Pred. No.: 6.43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-241-742-384 (1-346)

QY 1 ATGTCGAGGCCATGACGAAGTATCGCGGCGGCTGTCAGGCGCGGCTGTTCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATCTGACCGCACAGTCGGAAGACGATCGGCTGATCGACGAGCGGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGACTTGATTCGCGGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTyrLeu 60  
QY 181 GCGCGCGCGGCTTGGGCGATGCGCTTCTCCAGCGCTATTTCGAAGATTGCTGTCGCG 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAAGCAGTGGCAGGCGCTGCGGATGCGGCGCGCGCGCAGCATGATGTCGTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCAGGCGCGCGGCGGCGAGCGCTCTATATGAGCCAGCGCATCTTGGCGCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCATCTGATCGCCGCGCGCGGAGCTCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGGACCATCTCGCGGTGCACGATACCGGCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGGAGACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTCGCGCTGTCGGCGGAGCTTCAAGCTCTATCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGGAGCCAGATCTACGCGGTGCGAGGCGGCTGCTAC 660  
Db 201 LeuGlyProGluValAlaSerThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGGTATTGGTGATACGCC 720

Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleYsValLeuValAspThrPro 240  
QY 721 GACCAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGCCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCGCGCTCCGCGAGACCGAAGAGGAGTGTGTCGCCGATATGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGlyLeuLeuValAlaAspIleasp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAGCGCGCGCGCGATCCGGCGGGCCACTATTACGGCCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGGCGCTGCTGCTGATCGACGTCGGCGCCCAACCGCTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGGACAGGGCGACGCGCGCGCGCTGCGGTGGCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 4  
US-10-440-523-384  
; Sequence 384, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitric Oxide  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/440,523  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-384

Alignment Scores:  
Pred. No.: 6.43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0

Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0  
US-09-751-299-1 (1-1041) x US-10-440-523-384 (1-346)

QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGTGACGCGCGCGGTTCCTC 60  
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValAlaGlnAlaAlaProValPheLeu 20  
QY 61 GATCTCGACCCGACAGTCGAGAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATTTCCCGGCTATCCCTTTGGATATGGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTyrIleProGlyTyrProPheThrIleThrLeu 60  
QY 181 GCGCGCGCGCTTGGGCGATGCGCTTCTCGACGCGCTATTTCGAGAAATTGCTGCGC 240  
Db 61 GlyAlaProAlaTyrGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGACGTGCGCAGGCGCTGCGGATGCGCGCGCGCGCGCGCGCATGATGCTG 300  
Db 81 GlySerLysGlnTyrGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCGAGCGCGCGCGCGCGCGCGCTTATATGGCCAGGCGCATCTTGGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGACGACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGlnArgThrVal 140  
QY 421 TTCGGCGAGGAGAGCGGACCGCATCTCGCGGTGACAGATACCGCGCATGCGCGCTGCGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTGTGTTGGTGGAGACACATCCAGCCATTGTGAAATACGCCATGACCGCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTGCGCTGCGCGCGCGCGCTTACGCTTATCGCGCGCATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValAlaGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGTGCGCGGACCGCTTTCGCGCGAGATGATCAAGGTATTGGTGAATAAGCCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleYsValLeuValAspThrPro 240  
QY 721 GACCAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGCCGACGGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCCGAGCGCGCTCCGCGAGACCGAAGGAGGAGTGTGTCGCCGATATGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGlyLeuLeuValAlaAspIleasp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAGCGCGCGCGCGATCCGGCGGGCCACTATTACGGCCCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGGCGCTGCTGCTGATCGACGTCGGCGCCCAACGCTGTCACGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGGACAGGGCGACGCGCGCGCGCTGCGGTGGCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340



QY 1021 AGCGCCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaGln 346  
RESULT 5  
US-10-440-503-384  
; Sequence 384, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE  
; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440,503  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; PRIOR FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-384  
Alignment Scores:  
Pred. No.: 6.43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0  
US-09-751-299-1 (1-1041) x US-10-440-503-384 (1-346)  
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGTGCAGGCGCGGTGTTCTC 60  
Db 1 MetSerGluProMetThrLysTrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATCTGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGGCGCAAGCAGAC 120  
Db 21 AspleuAspargThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATCCCGGCTATCCCTTTGGATATGGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTriPleProGlyTyrProPheTriPleTriPleu 60  
QY 181 GCGCGCGCGCTGGGCGCATGCGCTTCGTCAGCGCTATTCGAGATTCGCTCGCGC 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGlnAsnSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGGCAGGCGCTGGCGGATGCGCGCGCGCGCACGCGCATGTGCTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgAlaGlnIleGlyMetHisValVal 100  
QY 301 GCCGCTATAGCAGCGCGCGCGCGCATCTATATGCGCAGCGCATCTTTCGCCCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCTATCGGAGCGCACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCAGGAGACGCGCAGCCATCTCGCGTGCACGATACCGCCATCGCGCGCTCGGC 480

Db 141 PheGlyGlnGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTGCTGGAGACACATCCAGCCATTGTCGAATACGCCATGTACGCCCGGAC 540  
Db 161 AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTrpAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCACGTGCGCTGTCGTCGCGCAGCTTCAGCCTCTATCGCGCATGCGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGCGCAGCCAGATCTACCGGTCGAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCGGAGATGATCAAGGTATTGCTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTTGCCATGATTTTCGCGCGCGCAGCGC 780  
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCTGCGCGCAGCGCGCTCCCGGAGACCGAAGAGGACTGCTGCTCGCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGCGATGATCGCGTTGGCCAAGCGCGCGCGATCCGCGCGGCACTATTCACGCGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTACCGCGCTGCTGCTGATGACGTCGCGCCCAACGCGTGTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValThrLeuAspAla 320  
QY 961 GCATTGCAACCGCAAAACGAGGACAGGCGCAGCGCGCGCTGCGGTGGCGGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValAlaGln 340  
QY 1021 AGCGCCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaGln 346  
RESULT 6  
US-10-461-925-384  
; Sequence 384, Application US/10461925  
; Publication No. US20040053378A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001  
; CURRENT APPLICATION NUMBER: US/10/461,925  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,317  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/392,944  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-384  
Alignment Scores:  
Pred. No.: 6.43e-123 Length: 346  
Score: 1806.00 Matches: 346



Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0  
US-09-751-299-1 (1-1041) x US-10-461-925-384 (1-346)

QY 1 ATGTCCGAGCCCATGACGAGATATCCGGCGCGCGGTGCAGGCGCGCGGTGTTCTC 60  
DB 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGCTGATCGAGCGGCGCCAGCAGGAC 120  
DB 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40  
QY 121 GTGGCGCTGATCGCATTCGCCAGAGACTTGATCCGGCTATCCCTTTGGATATGCTG 180  
DB 41 ValArgLeuIleAlaPheProGluThrTyrIleProGlyTyrProPheThrIleTrpLeu 60  
QY 181 GCGCGCGCGCTGGCGCATGCGCTTCGACGCGCTATTCGAATTCGATTCGCTCGCGC 240  
DB 61 GlyAlaProAlaTyrGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGGCAGGCGCGCTGGCGATGCGCGCGCGCCAGCGCATGCTGCTG 300  
DB 81 GlySerLysGlnTyrGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATAGCAGCGCGCGCGCGCGCTCTATATGGCGCCAGCGCATCTTCGCGCCC 360  
DB 101 AlaGlyTyrSerGluArgAlaGlyLysSerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCATCTGATCGCGCGCGCGCGCGCGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
DB 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGCGCAGGAGAGACGCGCAGCATTCGCGGTGCACGATACCGCATCGGCGCGCTCGGC 480  
DB 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTGCTGGAGACACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCGAC 540  
DB 161 AlaLeuCysCysTyrGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAAACAGTCCACGTGCGCTGCGCGCGCGCGCTTCAGCCTCTATCGCGCGCATGGCCTATGCG 600  
DB 181 GluGlnValHisValAlaSerTyrProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGACCGGAGGTCAATACCGCGCGCGCGCGGTTCGCATGATTTTCGGGCGCGACGGC 660  
DB 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGCTGCGCGCGCGCGCGCGGTTCGCGGAGATGATCAAGGTATTGGTATACGGCC 720  
DB 221 ValLeuAlaSerCysAlaThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAAGAGATGTTCTCAAGGCGCGCGCGGTTCGCATGATTTTCGGGCGCGACGGC 780  
DB 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCGCTGGCGGAGCGCGCTCCCGGAGACCGGAAGAGGAGTGTGCTGCGCGCATATCGAC 840  
DB 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluLeuLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATGCGGTTGGCCAAAGCGCGCGCGCATCGCGCGGCGCACTATTACGGCGCC 900  
DB 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGCGGCTGCTGCTGATCGACGTCCGCGCGCGCAAGCGCTGCACGCTGATGCC 960  
DB 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTGAAACGCAAAACGAGACAAAGGCGCAGCGCGCGCTGCGCGTGTGCGGAA 1020

DB 321 AlaPheGluProGlnAsnGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCGCGCAG 1038  
DB 341 SerAlaAlaAlaAlaGln 346  
RESULT 7  
US-10-146-772-34  
; Sequence 34, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
; US-10-146-772-34  
Alignment Scores:  
Pred. No.: 5.11e-98 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conserved: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 4 Gaps: 0  
US-09-751-299-1 (1-1041) x US-10-146-772-34 (1-341)  
QY 1 ATGTCCGAGCCCATGACGAGATATCCGGCGCGCGGTGCAGGCGCGCGGTGTTCTC 60  
DB 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGCTGATCGAGCGGCGCCAGCAGGAC 120  
DB 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGGCGCTGATCGCATTCGCCAGAGACTTGATCCGGCTATCCCTTTGGATATGCTG 180  
DB 41 ValArgLeuIleAlaPheProGluThrTyrIleProGlyTyrProLeuThrIleTrpLeu 60  
QY 181 GCGCGCGCGCTGGCGCATGCGCTTCGACGCGCTATTCGAATTCGATTCGCTCGCGC 240  
DB 61 GlySerProAlaTyrGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGGCAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCATGCTGCTG 300

Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGGCTATAGCGAGCGCGCGCGCGAGCCTTATATGAGCCAGGCGCATCTTGGCCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGAGCTCAAGCCTTACCCTAGCGGAGCGGACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGCAGCCATCTCCGGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTGCTGGAGACACATCCAGCATTGTGGAATACGCCCATGTACGCCCGCGAC 540  
Db 161 AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCGTGTGCGCGAGCTTCAAGCCTTATCGCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCySPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGCAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGTGCGCGACCGCTTCCGCGGAGATGATCAAGGTATTGGTATACGCC 720  
Db 221 ValLeuAlaSerCySerLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTGCCATGATTTTGGGCGCGACGCGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGCGCGCGCTCCCGGAGACCGAAGAGGACTGTGTGCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGGCATGATCGCGTGTGCGCAAGCGCGCGCGCGATCCGCGCGCGCACTATTCACGCGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTACCGCGCGCTGTGTGTGATCGAGCTCCGCGCGCAAGCGCTGTGTCACGCTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGCAACCGCAAAACGAGGACCAAGGCGCGCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 8  
US-10-241-742-34  
; Sequence 34, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-34

Alignment Scores:  
Pred. No.: 5,11e-98 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-241-742-34 (1-341)

QY 1 ATGTGCGAGCCCATGACGAGATTCGCGCGCGGCGGTGCAGGCGCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATTCGACCGCACAGTCGAGAAAGCATCGCGCTGATCGACGAGCGCGCAAGCAGGAC 120  
Db 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuTrpIleTyrLeu 60  
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAGAATTCCGCTGCGCG 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGTGCGAGCGCTGCGCGGATGCGCGCGCGCGCGCATGCTATGCTG 300  
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCCGGCTATAGCGAGCGCGCGCGCGAGCCTTATATGAGCCAGGCGCATCTTGGCCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGAGCTCAAGCCTTACCCTAGCGGAGCGGACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTCGGCGAGGAGACGGCAGCCATCTCCGGGTGCACGATACCGCCATCGGGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTGCTGGAGACACATCCAGCATTGTGGAATACGCCCATGTACGCCCGCGAC 540  
Db 161 AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCCACGTGCGCGTGTGCGCGAGCTTCAAGCCTTATCGCGCGCATGGCCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCySPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGACCGGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGCAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGTGCGCGACCGCTTCCGCGGAGATGATCAAGGTATTGGTATACGCC 720

Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValIleAspThrPro 240  
QY 721 GACAGAGAGATGTTCTCAAGGCGCGCGGTTTGGCATGATTTTGGGCGCGACGGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGCGAGCCGCTCCCGGAGACCGAAGAGGGACTGTGGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAGGCGCGCGCGCATCCGGCGGCGCACTATTCAAGCGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTAACGGCGCTGCTGTGATCGACGTCGCGCCCAACGCGTCGTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGGAACCGCAAAACGAGAGCAAGGCGCGACGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 9

US-10-440-523-34

; Sequence 34, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/440,523  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 34  
; LENGTH: 341  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-34

Alignment Scores:

Pred. No.: 5.11e-98 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 4 Gaps: 0

< US-09-751-299-1 (1-1041) x US-10-440-523-34 (1-341)

QY 1 ATGTCGAGACCCATGACGAAGTATCGCGCGCGCGGTGACGGCGCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGTATCGACGAGCGGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40  
QY 121 GTGGCCTGTATCGCATTCACAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCTTGGGCGATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGTCGCGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80  
QY 241 GCGAGCAAGCAGTGGCAGGCGCTTGGCGGATGCGCGCGCGCGCGCATGATGTCGTG 300  
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
QY 301 GCGGCTATAGCGAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGCGCATGTCGCGCC 360  
Db 101 ValGlyPheSerGlnArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGCGCATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCCTAACCATGCGGAGCGCACCGTG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGlnArgThrVal 140  
QY 421 TTCGGCGAGGAGAGACGGGAGCCATCTCGCGGTGCACGATACCGCCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160  
QY 481 GCGCTCTGTGCTGGGAGACATCCAGCCATTGTGAAATACGCCATGTACGCGCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCACGTCGCGCTGTGCGGAGCTTCAGCCTTATCGCGCATGCGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTCGGAACCGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTGCTGGCGTGTGCGCGACCGTTTCGCGCGAGATGATCAAGGTATTTGCGGATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAGAGAGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTTGGGCGCGACGGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTGGCGCGAGCGCTCCCGGAGACCGAAGAGGAGGACTGCTGTCGCCGATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTCGGCATGATCGCGTTGGCCAGGCGCGCGCGCATCCGGCGGCGCACTATTCAAGCGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrAlaArgPro 300  
QY 901 GACGTAACGGCGCTGCTGTGATCGACGTCGCGCCCAACGCGTCGTCAAGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGGAACCGCAAAACGAGAGCAAGGCGCGCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 10

US-10-440-503-34

; Sequence 34, Application US/10440503  
; Publication No. US20040038419A1

```

; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-34

Alignment Scores:
Pred. No.: 5.11e-98 Length: 341
Score: 1464.00 Matches: 272
Percent Similarity: 90.1% Conservative: 28
Best Local Similarity: 81.7% Mismatches: 33
Query Match: 77.8% Indels: 0
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-440-503-34 (1-341)

QY 1 ATGTCGAGCCCATGACGAGATATCCGCGCGCGGTGACAGCGCGCGGTTCCTC 60
   ||| ||:::||:::||::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MetLeuserProvalThrgIntYrArgAlaAlaValGlnAlaProSerPheLeu 20

QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCCAGCAGAC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21 AspleuaspargThrValGluLysThrIleAlaIleIleGluGlnAlaGluGlnAsp 40

QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTCGATTCCCGCTATCCCTTTGGATAGGCTG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuTrpIleTrpLeu 60

QY 181 GCGCGCGCGCGCTTGCGGCATCGCGCTTCGTCACGCCCTATTTCGAGAAATTCGCTCGTGCC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80

QY 241 GGCAGCAGCAGTCGAGCGCCCTGCGCGATGCGGCCCGCGCCAGCGCATGATGCTG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100

QY 301 GCCGCTATAGCGAGCGCGCGCGCGCGCTCTATATGCGCCAGCGCATTCGCGCCC 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120

QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGCGCAAGCTCAAGCCTACCCATCGCGAGCGCACCGTG 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140

QY 421 TTCGGCGAGGAGACGGCAGCGCCATTCGCGGTGCACGATACCGCCATCGCGCGCTCGGC 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrGluThrGlyValGlyArgGlyGly 160

QY 481 GCGCTCTGTTGCTGGAGACACATCCAGCCATTTGCGAAATACGCCATGTACGCGCGCAC 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180

QY 541 GAACAGGTCCACGTGCGCTGTGCGCGCGCGAGCTTCAGCCTCTATCGCGGCATGCGCTATGCG 600

```

```

Db      181  GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrrArgGlyMetAlaTyrrAla  200
QY      601  CTGGACCCGGAGGTCAATACCCGCCAAGCCAGATCTACGGCGTGCAGGGCGGTGCTAC  660
Db      201  LeuGlyProGluValaIaenThrrAlaAlaSerGlnValTyrrAlaValGluGlyCysTyrr  220
QY      661  GTGCTGGCGCTGTCGCCGACCCGTTTCGCCCGGAGATGATCAAGGTATTGGTGATACGCC  720
Db      221  ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro  240
QY      721  GACAAGAGATGTTCTCTCAAGCCGCCGCCGCGGCTTTGCCATGATTTTCGGGCCCGACGGC  780
Db      241  AspYrsGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly  260
QY      781  CGCGCCCTGGCCGAGCCCGCTCCCGGAGACCCGAAGAGGAGACTGCTGTGCCGATATCGAC  840
Db      261  ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp  280
QY      841  CTGGCGATGATCGCGCTTGCCCAAGCGCGCGCGCGATCCGGCGGCCACTATTTCACGGGCC  900
Db      281  LeuGlyAlaIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrrAlaArgPro  300
QY      901  GACGTAACGCGCGCTGCTGCTGATCGACGTCGCGCCCAACGCGTCGTCACGCTTGATGCC  960
Db      301  AspValThrArrLeuLeuLeuAsnProArrProAlaAlaArgValGluAlaLeuGlyPro  320
QY      961  GCATTGCAACCGCAAAACGAGACAGACAGCGCGCGCC  999
Db      321  ArgPheGluValValGlnSerGluGlnAlaGluProPro  333

RESULT 11
US-10-461-925-34
; Sequence 34, Application US//10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-34

Alignment Scores:
Pred. No.:          5.11e-98          Length:          341
Score:             1464.00           Matches:          272
Percent Similarity: 90.1%            Conservative:     28
Best Local Similarity: 81.7%          Mismatches:      33
Query Match:       77.8%             Indels:          0
DB:                4                 Gaps:            0

US-09-751-299-1 (1-1041) X US-10-461-925-34 (1-341)

QY      1  ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGCGGTGCAGGCCGCGCGGTTCCTC  60
Db      1  MetLeuSerProValThrGlnTyrArrAlaAlaAlaValGlnAlaAlaProSerPheLeu  20
QY      61  GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTTGATCGAGCAGCGCGCCCAAGCAGGAC  120

```



Db 21 AspleuAsparGThrValGluLysThrIleAlaIleIleGluGlnAlaIleGluGlnAsp 40  
Qy 121 GTGGCGCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTrpProLeuTrpIleTrpLeu 60  
Qy 181 GGGCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGC 240  
Db 61 GlycerProAlaTrpGlyMetArgPheValGlnArgTyrrPheGluAsnSerLeuValArg 80  
Qy 241 GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGCGCGCGCGCGCGCATGCGATGTCGTG 300  
Db 81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
Qy 301 GCCGGCTATAGCAGCGCGCGCGCGCGCGCTTATATGAGCCAGCGCATCTTCGGCCCC 360  
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrrMetGlyGlnAlaIlePheGlyPro 120  
Qy 361 GATGGCGATCTGATCGCGCGCGCGCGCGCGCTCAAGCCTTACCGCATCGCGAGCGCACCG 420  
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluArgThrVal 140  
Qy 421 TTCCGGCAGGAGAGCGCGAGCGCATCTTCGCGGTGACAGATACCGCGCATCGCGCGCTCG 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValTyrrGluThrGlyValGlyArgIleGly 160  
Qy 481 GCGCTGTGTTGCTGGAGACACATCCAGCGCATTTGCGAAATAGCGCATGTACGCGCGCG 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrrAlaMetTyrrAlaAlaAsn 180  
Qy 541 GAACAGATCCACGTGCGCGCTGCGCGCGCGCGCTTACGCTTATCGCGCGCATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrrArgGlyMetAlaTyrrAla 200  
Qy 601 CTCGGACCGGAGTCAATACCGCGCGCGCGCGCATGTACGCGGTGAGGGCGGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrrAlaValGluGlyCysTyrr 220  
Qy 661 GTGCTGGCGTCTGCGCGCGCGCGCTTTCGCGCGAGATGATCAAGTATTTGGTATACGCC 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
Qy 721 GACAAGGAGATGTTCTCAAGCGCGCGCGCGGTTTGGCATGATTTTCGGGGCCGACGGC 780  
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
Qy 781 CGCGCGCTGGCGCGCGCGCTCCCGGAGACCGAAGAGGAGTGTGCTGCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
Qy 841 CTCGGCATGATCGCGTTGGCCAAGCGCGCGCGCGCATCCGCGCGGCACTATTCACGGCCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrrAlaArgPro 300  
Qy 901 GACGTACCGCGGCTGCTGCTGATCGACGTCCCGGCCAACGCGTTCGTACCGCTTGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
Qy 961 GCATTCGAACCGCAAAACGAGACAAAGGCGACGCGGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 12

US-10-146-772-16  
; Sequence 16, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen

; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitriases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FaBSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 348  
; TYPE: PR  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-16

Alignment Scores:  
Pred. No.: 1.89e-79 length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 4 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-146-772-16 (1-348)

Qy 10 CCCATGACGAG--TATCCGCGCGCGCGGTGACAGCCCGCGGTGTTCTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
Qy 67 GACCGCACAGTCGAGAAACCGATCGCGCTGATTCGACGACGCGGCCAACGACGACGTGCC 126  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaAlaSerAsnGlyAlaLys 41  
Qy 127 CTGATTCGATTCGCCAGACATTGGATTCCCGCGCTATCCCTTTTGATATGCTGCGCGCG 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrrProTrpTrpIleTrpLeuAspSer 61  
Qy 187 CCGGCTTGGGCGATGCGCTTCGTCCAGCGCTATTTGAGAAATTCGCTCGTCCGCGGACG 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrrPheAspAsnSerLeuMetLeuGlySer 81  
Qy 247 AAGCAGTGGCAGGCGCTGCGCGCGATGCGGCGCGCGCGCGCATGATGTCGTGCGCGCG 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrrValValMetGly 101  
Qy 307 TATAGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCATCTTCGCGCGCGATGCC 366  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrrMetGlyGlnSerIleIleAsnAspLysGly 121  
Qy 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTTACCGCATCGGAGCGCACCGGTTCGGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly 141  
Qy 427 GAGGAGACGCGCAGCGCATCTCGCGGTGACAGTACCGCGCATCGGCGCGCTCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCysValMetAspThrGluIleGlyArgValGlyAlaMet 161  
Qy 487 TGTGCTGGAGACATCCAGCCATTTGGAATAAGCCATGTACGCGCGCGCGCAAGAACAG 546

Db 162 CysCyvTtrpGlnHisLeuGlnProLeuSerLysTyraIaMetTyrsErgIaAspGluGln 181  
QY 547 GTCCACGTGCGCGTGGCGGAGCTTCAGCCTCTATGCGGCGATGCGCTATGCGCTGGA 606  
Db 182 ILeHisIleAlaSerTrpProSerPheSerLeuTyraGlyAlaIaIaTyraIaLeuGly 201  
QY 607 CCGGAGGTCAATACCGCGCGCAAGCCAGATCTACCGCGTGAAGGCGCGGTGCTAGCTGCTG 666  
Db 202 ProGluLeuAsnAsnAlaIaIaSerGlnMetTyraIaIaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCGTGCGCGACCGCTTTCGCGGAGATGATCAAGGTATTGGTGATACCGCGCAAG 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProAlaGly 241  
QY 727 GAGATGTTCTCAAGGCGCGCGCGGCTTTGGCCATGATTTTCGGCGCGCGCGCGCGCC 786  
Db 242 GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyraGlyProAspGlyArgPro 261  
QY 787 CTGCGCGAGCGCGTCCCGGAGACCGAAGAGGAGTGTGTCGCCGATATCGACTTCGCGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyraIaAspIleAspLeuGly 281  
QY 847 ATGATCGCGTTGGCCCAAGCGCGCGCGCGATCCGCGCGGCGCACTATTCAACGCGCGCGTA 906  
Db 282 MetIleSerMetAlaLysAlaIaIaAspProAlaGlyHisTyraIaArgProAspVal 301  
QY 907 ACGCGGCTGCTGCTGATCGACGTCCGCGCGCAACGCGCTGTACGCTTGATGCGCGCATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyriIeAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGCAAGGCGGACGCGCGCGCGCTGCGCTGTGTCGCGCAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluIaGluGlnGln 340  
QY 1027 GCCGCGCGC 1035  
Db 341 LysAlaIaIa 343  
RESULT 13  
US-10-241-742-16  
; Sequence 16, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitricases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 348

; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-16  
Alignment Scores:  
Pred. No.: 1.89e-79 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 4 Gaps: 2  
US-09-751-299-1 (1-1041) x US-10-241-742-16 (1-348)  
QY 10 CCCATGACGAAG--TATCGCGCGCGCGGTGCAGCGCGCGCGGTGTTCTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaIaValGlnAlaIaIaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCCAGAAAGCGATCGCGCTGATCGAGCAGCGCGCGCAAGCAGACGTCGCGC 126  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaIaIaIaSerAsnGlyAlaLys 41  
QY 127 CTGATCGCATTCACAGACTTGATTCGCGGCTATCCCTTTGGATATGCGTGGCGCGC 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyroProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGGATGCGCTTGTTCAGCGCTATTTCAGAAATTGCTGTCGCGCGCAGC 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyroPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGCAGGCCCTGCGCGATGCGCGCGCGCGCGCACGCGCATGCTGTGCGCGC 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaIaIaAsnAsnLysIleTyroValValMetGly 101  
QY 307 TATAGCAGCGCGCGCGCGCGAGCCTTATAGGCCAGCGCATCTTCGCGCGCGATGCGC 366  
Db 102 TyrsErgGluArgSerGlyGlySerLeuTyroMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGTGTGCGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGlnArgThrValPheGly 141  
QY 427 GAGGAGAGCGGAGCCATCTCGCGGTGACGATACCGGCATCGGGCGCTGCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCysValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTTCGTGGAGCACATCCAGCCATTTGCAATAACGCCATGTACGCGCGCGCAAGCAG 546  
Db 162 CysCyvTtrpGlnHisLeuGlnProLeuSerLysTyraIaMetTyrsErgIaAspGluGln 181  
QY 547 GTCCACGTGCGCGTGGCGGAGCTTCAGCCTCTATGCGCGCGCATGCGCTATGCGCTCGA 606  
Db 182 ILeHisIleAlaSerTrpProSerPheSerLeuTyraGlyAlaIaIaTyraIaLeuGly 201  
QY 607 CCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTGCTAGCTGCTG 666  
Db 202 ProGluLeuAsnAsnAlaIaIaSerGlnMetTyraIaIaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCGTGCGCGACCGTTTTCGCGGAGATGATCAAGGTATTGGTGATACCGCGCAAG 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGGCGCGCGCGGCTTTGGCCATGATTTTCGGCGCGCGCGCGCGCC 786  
Db 242 GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyraGlyProAspGlyArgPro 261  
QY 787 CTGCGCGAGCGCGTCCCGGAGACCGAAGAGGAGTGTGTCGCCGATATCGACTTCGCGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyraIaAspIleAspLeuGly 281  
QY 847 ATGATCGCGTTGGCCCAAGCGCGCGCGCGATCCGCGCGCGCACTATTCAACGCGCGCGTA 906

Db 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAlaSpVal 301  
QY 907 ACGCGCTGCTGTGATCGACGTCCGCGCCCAACCGCTGTCACGCTTGATGCCGATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGACACAGGGCGACGCGCCCGCTGCGCGCTGTGGCGCAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGlnGln 340  
QY 1027 GCCGCGCGC 1035  
Db 341 LysAlaAla 343

RESULT 14  
US-10-440-523-16

; Sequence 16, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/440,523  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
; US-10-440-523-16

Alignment Scores:  
Pred. No.: 1.89e-79 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 4 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-440-523-16 (1-348)

QY 10 CCCATGACGAAG--TATCGCGCGCGCGCGGTGCAGGCGCGCGCGGTTCCTCGATCTC 66  
Db 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCGAGAAAGCGATCGGCTTGATCGAGCAGCGCGCGCCCAACGAGACGATGCCG 126  
Db 22 GluGlyAlaIleSerLysGlyIleSerLeuIleGluGluAlaAlaSerAsnGlyAlaLys 41

QY 127 CTGATCGCATTCACAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCGTGGCGCG 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGACATGCGCTTCGTCACGGCTATTTCGAGAATTGCTGCGCGGACG 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGGCCCTGGCGGATGCGGCGCGCCGCGCACGCGCATGATGTCGCGCGCG 306  
Db 82 GluGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101  
QY 307 TATAGCAGCGCGCGCGCGGCGAGCCTCTATATGCGCCAGCGCATCTTCGCGCCGATGCG 366  
Db 102 TyrSerGluArgSerGlyGlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCCATGCGGACGCGATGTTCCGC 426  
Db 122 GluThrIlePheThrArgArgLysLeuLysProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGGGACGACCATCTCGCGGTGCACGATACCGGCATCGGCGCGCTCGCGGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCyValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTGTCTGGAGACACATCCAGCCATGTGCAAAATACGCCATGTACGCCCGACGAACAG 546  
Db 162 CysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGln 181  
QY 547 GTCCACGTGCGCTGCTGTGGCGCGAGCTTCAGCCTCTATCGCGCATGCGCTTGCCTCGA 606  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGGAGTCAATACCGCGCGCAAGCCAGATCTACGGGTGAGGCGCGCTCTACGTGCTG 666  
Db 202 ProGluLeuAsnAsnAlaAlaAspGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCTGCGCGACCGTTTCGCGCGGATGATCAAGGTATGATGATACGCCGCGCGC 726  
Db 222 AlaProCysAlaThrValSerLysGluMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGGCGCGCGGCTTTTGCATGATTTTCGGGCGCGACGCGCGC 786  
Db 242 GluProLeuLeuLeuGluGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGCGCGAGCGCGCTCCCGGAGACGGAAGGAGACTGTGTCGCGCATATCGACCTCGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGGCTTGCCCAAGGCGCGCGCGATCCGCGGCGCCACTATTCACGCGCGCAGCTA 906  
Db 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301  
QY 907 ACGCGCTGCTGTGATCGACGTCGCGGCCCAACGCGCTGTCACGCTTGATGCCGATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGACACAGGGCGACGCGCGCGCTGCGCGTGTGGCGCAAGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGlnGln 340  
QY 1027 GCCGCGCGC 1035  
Db 341 LysAlaAla 343

RESULT 15  
US-10-440-503-16

; Sequence 16, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen

APPLICANT: Milan, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Burk, Mark J.  
APPLICANT: McQuaid, Jeffrey  
APPLICANT: Stege, Justin  
TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
NITRILES AND/OR CYANIDE  
FILE REFERENCE: 09010-900001  
CURRENT APPLICATION NUMBER: US/10/440,503  
PRIOR APPLICATION NUMBER: 2003-05-15  
PRIOR FILING DATE: 2002-05-15  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 16  
LENGTH: 348  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-16

Alignment Scores:  
Pred. No.: 1.89e-79 length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
DB: 4 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-440-503-16 (1-348)

QY 10 CCCATGACGAG--TATCGCGCGCGCGGTGTCAGCGCGCGCGGTTCCTCGATCTC 66  
Db 2 ProThrSerIysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCCGACAGTCGAGAAAGCATCGGCTGATCGACGAGCGCGCCAGCAGCAGCTGCC 126  
Db 22 GluGlyAlaIleSerIysGlyIleSerLeuIleGluGlnAlaAlaSerAsnGlyAlaLys 41  
QY 127 CTGATCGCATTCACAGACTTGGATTCCCGGCTATCCCTTTTGATATGCTGGCGCG 186  
Db 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61  
QY 187 CCGGCTTGGGCGATGCGCTTCCTCCAGCGCTATTTCGAGAAATTGCTCGCGCGCAGC 246  
Db 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGCGCCCTGGCGGATGCGGCCCGCGCCGACGCGCATGTCGTGCGCGGC 306  
Db 82 GluGlnAlaIysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101  
QY 307 TATAGCGAGCGCGCGCGCGCGAGCTCTATATGCGCGCGCGCATCTTCGCGCCGATGGC 366  
Db 102 TyrSerGluArgSerGlyIleSerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121  
QY 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGAGCGCACCCTGTTCCGC 426  
Db 122 GluThrIlePheThrArgValGlyLeuLysProThrHisValGluArgThrValPheGly 141  
QY 427 GAGGAGACGCGCAGCCATCTCGCGGTGCAGATACCGCATCGGCGCGCTTCGCGCGCTC 486  
Db 142 GluGlyAspGlySerHisLeuCysValMetAspThrGluIleGlyArgValGlyAlaMet 161  
QY 487 TGTTCGTCGAGACACATCCAGCCATTGTGAATATACGCATGTACGCCGCGCAAGACAG 546  
Db 162 CysCysTrpGluHisLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGln 181  
QY 547 GTCCACGTCGCGTGTGGCGGAGCTTACGCTTATCGCGCGCATGGCCTATGCGCTCGGA 606  
Db 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaTyrAlaLeuGly 201  
QY 607 CCGAGGTCAATACCGCGCGCAAGCAGATCTACGCGGTGAGGGCGGCTGCTACGTGCTG 666

Db 202 ProGluLeuAsnAsnAlaAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeu 221  
QY 667 GCGTCGCGCGACCGTTTCGCGCGAGATGATCAAGTATTGGTGATACGCCGACAG 726  
Db 222 AlaProCysAlaThrValSerIleSerLysGlnMetIleGluMetLeuIleAspAspProArgLys 241  
QY 727 GAGATGTTCTCAAGCGCGCGCGGTTTTCATGATTTCGGGCGCGCGCGCGCGCC 786  
Db 242 GluProLeuLeuGluGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGCGCGAGCGCTCCCGGAGACCGGAGGGACTGCTGTCGCGCATATCGACCTCGGC 846  
Db 262 LeuAlaLysProLeuProGluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGCGTGGCCCAAGCGCGCGCGCATCCGCGCGCGCGCGCGCGCGCGCGTA 906  
Db 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301  
QY 907 ACGCGCTGCTGCTGATCGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATTC 966  
Db 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGluTyrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAAGCAGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 1026  
Db 322 GlyPro--ThrGluSerLeuLysAspMetGlyLysMetGlnMetGluAlaGluGln 340  
QY 1027 GCCGCGCGCG 1035  
Db 341 LysAlaAla 343

Search completed: April 27, 2006, 01:51:41  
Job time : 172.128 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:41:06 ; Search time 4.05255 Seconds  
(without alignments)  
2336.425 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1881  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgcaagtag 1041

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 232119 segs, 45477862 residues

Total number of hits satisfying chosen parameters: 464238

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/abs/ABSSWEB/spool/US09751299/runat\_26042006\_090044\_18835/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_New -QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US09751299 @CCGN 1 1 47 @runat\_26042006\_090044\_18835 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA\_New:  
1: /SIDS5/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /SIDS5/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /SIDS5/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /SIDS5/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
5: /SIDS5/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
6: /SIDS5/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /SIDS5/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /SIDS5/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	869	46.2	356	6	US-10-537-075-7	Sequence 7, Appli
2	810	43.1	369	6	US-10-919-182-16	Sequence 16, Appl
3	809	43.0	369	6	US-10-919-182-18	Sequence 18, Appl
4	807	42.9	369	6	US-10-919-182-4	Sequence 4, Appli
5	807	42.9	369	6	US-10-919-182-6	Sequence 6, Appli
6	805	42.8	369	6	US-10-919-182-12	Sequence 12, Appl
7	804	42.7	369	6	US-10-919-182-14	Sequence 14, Appl
8	796	42.3	369	6	US-10-919-182-8	Sequence 8, Appli
9	430.5	22.9	333	7	US-11-096-568A-20687	Sequence 20687, A

10	430.5	22.9	351	7	US-11-096-568A-20686	Sequence 20686, A
11	399.5	21.2	330	7	US-11-096-568A-6955	Sequence 6955, Ap
12	399.5	21.2	350	7	US-11-096-568A-6954	Sequence 6954, Ap
13	315	16.7	233	7	US-11-096-568A-6956	Sequence 6956, Ap
14	284	15.1	193	7	US-11-096-568A-20688	Sequence 20688, A
15	241	12.8	449	7	US-11-096-568A-20255	Sequence 20255, A
16	240.5	12.8	532	7	US-11-096-568A-22026	Sequence 22026, A
17	240.5	12.8	702	7	US-11-096-568A-14646	Sequence 14646, A
18	213	11.3	352	7	US-11-096-568A-23018	Sequence 23018, A
19	206.5	11.0	438	7	US-11-096-568A-16134	Sequence 16134, A
20	206	11.0	392	7	US-11-096-568A-21010	Sequence 21010, A
21	205	10.9	438	7	US-11-096-568A-16134	Sequence 16134, A
22	205	10.9	512	7	US-11-096-568A-25890	Sequence 25890, A
23	205	10.9	535	7	US-11-096-568A-25889	Sequence 25889, A
24	204.5	10.9	379	7	US-11-096-568A-17169	Sequence 17169, A
25	203.5	10.8	312	7	US-11-096-568A-9235	Sequence 9235, Ap
26	202.5	10.8	485	7	US-11-096-568A-11820	Sequence 11820, A
27	201.5	10.7	401	7	US-11-096-568A-9081	Sequence 9081, Ap
28	200	10.6	464	7	US-11-096-568A-12965	Sequence 12965, A
29	199	10.6	413	7	US-11-096-568A-27265	Sequence 27265, A
30	199	10.6	442	7	US-11-096-568A-19628	Sequence 19628, A
31	198.5	10.6	286	7	US-11-096-568A-6597	Sequence 6597, Ap
32	198.5	10.6	422	7	US-11-096-568A-6597	Sequence 6597, Ap
33	198.5	10.6	458	7	US-11-096-568A-6596	Sequence 6596, Ap
34	198	10.5	273	7	US-11-096-568A-24596	Sequence 24596, A
35	197.5	10.5	375	7	US-11-096-568A-23618	Sequence 23618, A
36	197	10.5	424	7	US-11-096-568A-22088	Sequence 22088, A
37	196	10.4	758	7	US-11-096-568A-14593	Sequence 14593, A
38	196	10.4	766	7	US-11-096-568A-19864	Sequence 19864, A
39	196	10.4	1229	7	US-11-188-298-7436	Sequence 7436, Ap
40	195	10.4	333	7	US-11-096-568A-19235	Sequence 19235, A
41	193	10.3	276	7	US-11-096-568A-16388	Sequence 16388, A
42	192.5	10.2	429	7	US-11-096-568A-20326	Sequence 20326, A
43	192	10.2	386	7	US-11-096-568A-21828	Sequence 21828, A
44	190.5	10.1	386	7	US-11-096-568A-21828	Sequence 21828, A
45	190	10.1	276	7	US-11-096-568A-16388	Sequence 16388, A

ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zelinski, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-10-537-075-7

Alignment Scores:  
Pred. No.: 1,31e-52 Length: 356  
Score: 869.00 Matches: 171  
Percent Similarity: 66.2% Conservative: 54  
Best Local Similarity: 50.3% Mismatches: 105  
Query Match: 46.2% Indels: 10  
DB: 6 Gaps: 2  
US-09-751-299-1 (1-1041) x US-10-537-075-7 (1-356)

QY	25	CGCGGCGCGCGGTGCAGGCGCGCGCGGTTCCTCGAATCTGCACCGCACAGTCGAGAA	84
Db	8	ArgAlaAlaAlaValGlnAlaAlaSerProAsnTyraSpleuAlaThrGlyValAspLys	27
QY	85	GGCATCGGCGTGCATCGAGCAGCGCGCCAAAGCAGACGTGCGCTGATCGCATTCAGAG	144
Db	28	ThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCysAspLeuIleValPheGlyGlu	47
QY	145	ACTTGATTCGCGGCTATCCCTTTGGATATAGCTGGGCGCGCGCTTGAGGCAATCGC	204
Db	48	ThrTrpLeuProGlyTyrProPheHisValTrpLeuGlyAlaProAlaTrpSerLeuLys	67
QY	205	TTGCTCCAGCGCTATTTCGAGAATTCGCTCGTCCGCGGCGAGCAAGCAGTGGCAGGCC	264
Db	68	TyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGluPheGlnArgIle	87
QY	265	GCGGATGCGCGCGCGCGCGCGCATGCATGCTGCGCGCGCTATAGCAGCGCGGCG	324
Db	88	AlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSerGluArgSerGly	107
QY	325	GGCAGCGCTTATATGGGCCAGCGCATCTTCGCGCGCGCATGGCGATCTGATCGCGCGC	384
Db	108	GlySerLeuTyrLeuGlyGlnCysLeuIleAspAspLysGlyGluMetLeuTrpSerArg	127
QY	385	CGCAAGCTCAAGCCTTACCCTATCGGAGCGCGCACCGTGTTCCGCGAGGAGCGAGCC	444
Db	128	ArgLysLeuLysProThrHisValGluArgThrValPheGlyGlyTyrAlaArgAsp	147
QY	445	CTCGCGGTGCACGATACCGCCATCGGCGCGCGCTGCGCGCTGTGTTGCTGGAGCACATC	504
Db	148	LeuIleValSerAspThrGluLeuGlyArgValGlyAlaLeuCysCysTrpGluHisLeu	167
QY	505	CAGCCATTGTCGAATAACGCCATGTACGCCCGCGCGAGCAACAGGTCCACGTCGCGTGTG	564
Db	168	SerProLeuSerLysTyrAlaLeuTyrSerGlnHisGluAlaIleHisIleAlaIleTrp	187
QY	565	CCGAGCTTCAGCCTCTATCGCGCGCATGGCCTATGCGCTCGAGCCGAGGTCAATACGCC	624
Db	188	ProSerPheSerLeuTyrSerGluGlnAlaHisAlaLeuSerAlaLysValAsnMetAla	207
QY	625	GCAAGCCAGATCTACGCGGTGCAGGCGCGCTGCTACGTGCGCTGCTGCGCGCACCGTT	684
Db	208	AlaSerGlnIleTyrSerValGlyGlyGlnCysPheThrIleAlaIleSerSerValVal	227
QY	685	TCGCGCGAGATGATCAAGGTATTGGTGATACGCCCGACAAAGAGATGTTCTCAAGGCC	744
Db	228	ThrGlnGluThrLeuAspMetLeuGluValGlyGluHisAsnAlaProLeuLeuLysVal	247
QY	745	GGCGGCGGTTTGCCATGATTTTCGGGCGCGCGCGCGCGCGCTGCGCGAGCCGCTCCG	804
Db	248	GlyGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAlaProTyrLeuPro	267
QY	805	GAGACCCGAAGAGGACTGCTGTCGCGCGATATCGACCTCGGCGATGATCGCGTGGCCAAG	864
Db	268	HisAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluGluIleAlaPheAlaLys	287
QY	865	GCGGCGCGCGATCCGCGCGGCGCACTATTCACGCGCGCGACGTAACGCGGCTGCTGGAT	924
Db	288	AlaIleAsnAspProValGlyHisTyrSerLysProGluAlaThrArgLeuValLeuAsp	307
QY	925	CGACGTCCGCGCCCAACGCGTCTCACGCTTGATGCCGATTGGAACCGCAA	975
Db	308	LeuGlyHisArgAspProMetThrArgVal-----HisSerLys	320
QY	976	AACGAGACAAGGCGACGCGCGCGCTGCGCGTGTGCGCGAAACGCGCGCGCGCG	1035
Db	321	SerValThrArgGluGluAlaProGluGlnGlyValGlnSerLysIleAlaSerValAla	340

```

: GENERAL INFORMATION:
: APPLICANT: E.I. duPont de Nemours and Company, Inc.
: APPLICANT: Di Cosimo, Robert
: APPLICANT: Payne, Mark
: APPLICANT: O'Keefe, Daniel
: TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
: FILE REFERENCE: CL2584 US NA
: CURRENT APPLICATION NUMBER: US/10/919,182
: CURRENT FILING DATE: 2004-08-16
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 16
: LENGTH: 369
: TYPE: PRT
: ORGANISM: Artificial sequence
: FEATURE:
: OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change
US-10-919-182-16

Alignment Scores:
Pred. No.: 1.53e-48 Length: 369
Score: 810.00 Matches: 160
Percent Similarity: 62.1% Conservative: 53
Best Local Similarity: 46.6% Mismatches: 126
Query Match: 43.1% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-16 (1-369)

QY 16 ACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTCTTCGTGATTCGACCGCACA 75
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6 SerlySphenleuAlaIaThrValGlnAlaGluProValTrpLeuAspAlaSpAlaThr 25

QY 76 GTCGAGAAAGCCATCGCGCTGATCGACGAGCGCGCCCAAGACGAGACGTCGCGCTGATCGCA 135
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 26 IleAsplySerIleGlyIleIleGluGluAlaIaGlnlyGlyAlaSerIleAla 45

QY 136 TTCCAGAGACTTGATTCGGGCTATCCCTTTTGATATGCTGGCGCGCGCTTG 195
   ||||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 46 PheProGluValPheIleProGlyTyrProTyrTrpAlaTrpLeuGlyAspVallySTyr 65

QY 196 GGCATGCGCTTCGTCAGCGCTATTTCGAATTCGCTCGTGGCGCGCAGCAAGCATGG 255
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85

QY 256 CAGGCCCTGGCGGATGCGCGCGCGCCGACCGCATGCATGTCGTGGCGCGCTATAGCGAG 315
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 86 ArgArgLeuGlnLeuAlaIaArgArgAsnlySileAlaLeuValMetGlyTyrSerGlu 105

QY 316 CGCGCGCGCGCAGCCTCTATATGGGCCAGCGCATCTTCGGCCCCGATGGCGATCTGATC 375
   || ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125

QY 376 GCCGCGCGCGCAAGCTCAAGCCTACCATGCGGACGCGACCGCTGTTCGGCGAGGAGAC 435
   || ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 126 AlaAsnArgArglySleuLySProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145

QY 436 GGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCGCTCGCGCGCTCTGTGCTGG 495
   || ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCySTrp 165

QY 496 GAGCACATCCAGCATTTGTGGAATACGCCATGTACGCGCGCGCAAGACAGGTCCACGTC 555
   || ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 GluHisValGlnProLeuSerLySPhaMetMetTyrSerLeuGlyGluGlnValHisVal 185

QY 556 GCGTGTGCGCGAGCTTCAGCCTCTATCGCGCGCATGACCTATGCGCTCGGACCGAGGTC 615
   || ||| ||||| ::|| ||| ::|| ||| ::|| ||| ||| ||| |||
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205

QY 616 AATACCGCGCAAGCCAGATCTACCGCGGTGAGGCGCGCTGCTACGCTGCGTGTGC 675
   || ||| ::::: ||||| ::||| ||| ||| ||| ||| ||| ||| |||
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCySerThr 225

```

```
QY 676 GCGACCGTTTCGCGGAGATGATCAAGGTATTGTGATACGCCGACAGAAGATGTTTC 735
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245
QY 736 CTCGAAGCCGCGCGGTTTGGCATGATTTTCGCGCCGACGCGCGCCCTGGCCGAG 795
Db 246 LeuProGlnGlyCysGlyTrrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265
QY 796 CCGCTCCCGGAGACCGAAGAGGACTGCTGTGTCGCCGATATCGACCTCGGCATGATCGCG 855
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285
QY 856 TTGGCCAAAGCGCGCGCGCATCCGCGCGGCACTATTCACGCGCCGACGTAACGCGGCTG 915
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305
QY 916 CTGCTGATCGACGT-----CCGCCCCAACGCGTTCGTACGCTTGATGCGCGCATTC 966
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324
QY 967 GAACCGCAAAACGAGACAGAGGCGACGCGCGCGCGCTGCGCGTGTGCGGAAAGCGCC 1026
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaAlaGluGlnGlu 344
QY 1027 GCCCGCCGCG 1035
Db 345 ArgGlnAla 347

RESULT 3
US-10-919-182-18
; Sequence 18, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change
US-10-919-182-18

Alignment Scores:
Pred. No.: 1.8e-48 Length: 369
Score: 809.00 Matches: 160
Percent Similarity: 62.1% Conservative: 53
Best Local Similarity: 46.6% Mismatches: 126
Query Match: 43.0% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-18 (1-369)
QY 16 ACGAAGTATCGCGCGCGGTGACAGCGCGCGGTGTTCTCGATCTGACCGCAC 75
Db 6 SerLysPheLeuAlaIleThrValGlnAlaGluProValTrrPleuAspAlaAspAlaThr 25
QY 76 GTCGAGAAGCGATCGCGCTGATCGAGCAGCGCGCGCAAGCAGAGACGTGCGCTGATCGCA 135
Db 26 IleAspLysSerIleGlyIleIleGluGluAlaGlnLysGlyAlaSerLeuIleAla 45
QY 136 TTCCCAAGACTTGATCCCGGCTATCCCTTTTGATATGCTGGCGCGCGGCTTGG 195
Db 46 PheProGluValPheIleProGlyTyrProTyrTrrPalaTrrPleuGlyAspValLysTyr 65
```

```
QY 196 GGCATGCGCTTCGTTCCAGCGCTATTTCGAGATTTCGCTGTCGCGGACGACGAGTGC 255
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85
QY 256 CAGGCCCTGGCGGATGCGGCGCGCGCGCGCGCATGATGTCGTGGCGCGCTATAGCGAG 315
Db 86 ArgArgLeuGlnLeuAlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105
QY 316 CCGCGCGGCGGACCTCTATATGCGGCGGCGCATCTTCGCGCGCGCATCTGATC 375
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125
QY 376 GCCGCGCGCGCAAGCTCAAGCTTACCGCGGAGCGCACCGTTCGCGGAGGAGAC 435
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145
QY 436 GGCAGCATCTCGCGGTGCACGATACCGGCAATCGGCGCGCTCGCGCGCTGTGCTGG 495
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyLysLeuAsnCysTrr 165
QY 496 GAGCACAATCCAGCATTTGTCGAATACGCCATGTACGCCGCGCGGACGAGAGTCCAGTCC 555
Db 166 GluHisLeuGlnProLeuSerLysPheMetTyrSerLeuGlyGluGlnValHisVal 185
QY 556 GCGTCTGCGCGGACTTCAGCCTCTATCGCGGCAATGCGCTATGCGCTCGGACCGGAGTCC 615
Db 186 AlaSerTrrProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205
QY 616 AATACCGCGCGGACCGAGATCTACGCGGTGAGGCGCGCTGCTACGTGCGCTGCTGTC 675
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGluGlnThrPheValLeuCysSerThr 225
QY 676 GCGACCGTTTCGCGGAGATGATCAAGGTATTGTGATACGCCGACAGAAGATGTTTC 735
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245
QY 736 CTCGAAGCCGCGCGGTTTGGCATGATTTTCGCGCCGACGCGCGCCCTGGCCGAG 795
Db 246 LeuProGlnGlyCysGlyTrrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265
QY 796 CCGCTCCCGGAGACCGAAGAGGACTGCTGTGTCGCCGATATCGACCTCGGCATGATCGCG 855
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285
QY 856 TTGGCCAAAGCGCGCGCGCATCCGCGCGGCACTATTCACGCGCGCGCATGACGCGGCTG 915
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305
QY 916 CTGCTGATCGACGT-----CCGCCCCAACGCGTTCGTACGCTTGATGCGCGCATTC 966
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324
QY 967 GAACCGCAAAACGAGACAGAGGCGCGCGCGCGCGCTGCGCGTGTGCGGAAAGCGCC 1026
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaAlaGluGlnGlu 344
QY 1027 GCCCGCCGCG 1035
Db 345 ArgGlnAla 347

RESULT 4
US-10-919-182-4
; Sequence 4, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
```



```
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Acidovorax facilis 72M
US-10-919-182-4

Alignment Scores:
Pred. No.: 2.47e-48 Length: 369
Score: 807.00 Matches: 160
Percent Similarity: 61.8% Conservative: 52
Best Local Similarity: 46.6% Mismatches: 127
Query Match: 42.9% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-4 (1-369)

QY 16 ACGAAGTATCGCGCGCGGTGCGAGCCCGCGGTGTTCTCGATCTCGACCGCAC 75
   :::::::::: ||| ||||| ||||| ::||| ||| |||
Db 6 SerLySpHeLeuAlaIaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThr 25

QY 76 GTCGAGAAAGCGATCGGCCTGATCGAGCGCGGCCAAGCAGCAGCAGCGCTGATCGCA 135
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 26 IleAspLySerIleGlyIleIleGluGlnAlaIaGlnLySGLyAlaSerLeuIleAla 45

QY 136 TTCCCGAGACTTGAGATTCCCGGCTATCCCTTTTGATATGCTGGCGCGCGCTTG 195
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 46 PheProGluValPheIleProGlyTrpProTyTrpAlaTrpLeuGlyAspValLysTyr 65

QY 196 GGCATGCGCTTCGTCAGCGCTATTTCGAGAAATTCGCTCGCGCGCGCAGCAAGCAGTGG 255
   ::||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85

QY 256 CAGGCCCTGGCGGATCGGCGCGGCCCGCCAGCATGTCGTCGCGCGCTATAGCGAG 315
   ::||| ||| ||||| ||||| ||||| ::||| ::||| ||||| ||||| |||||
Db 86 ArgArgLeuGlnLeuAlaIaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105

QY 316 CGCGCGCGCGCAGCCTTATATGGCCAGCGCATCTTCGCCCCGATGGCGATCGATC 375
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125

QY 376 GCCGCGCGCGCAAGCTCAAGCCTTACCCATGCGGAGCGCACCGCTGTGGCGAGGAGAC 435
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145

QY 436 GCGACGCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGCGCGCTCTGTGCTGG 495
   ||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrp 165

QY 496 GAGCACATCCAGCCATTTGTGGAATACGCCATGTACGCCCGCGCAACAGTCCACGTC 555
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 166 GluHisPheGlnProLeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisVal 185

QY 556 GCGTCGTGCGCGAGCTTCAGCCTCTATCGCGCGCATGGCCTATGCCGTCGGAACCGAGGTC 615
   ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205

QY 616 AATACCGCGCGAAGCCAGATCTACGCGGTGAGGGCGGCTGTACGTGCTGGCTGTC 675
   ||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGlyGlnThrPheValLeuCysSerThr 225

QY 676 GCGACCGTTTCGCGGAGATGATCAAGTATTGGTGATACGCCCGCAAGAGAGATGTC 735
   ::||| ||| ||||| ||||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245

QY 736 CTCAAGCGCGCGCGGCTTTGCGCATGATTTTCGCGCGCGCGCGCGCGCTGCGCGAG 795
   ||| ||| ||||| ||||| ::||| ::||| ||||| ||||| ||||| ||||| |||||
Db 246 LeuProGlnGlyCysGlyTrpAlaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265

QY 796 CCGCTCCGCGAGACCGAAGAGGAGCTGTGTCGCGCATATCGACCTCGGCATGATCGCG 855
   ||||| ||||| ||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285
```

```
QY 856 TTGGCCAAAGCGCGCGCGCATCCGGCGGCCACTATTTCACGCGCCGACGTAAAGCGGCTG 915
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305

QY 916 CTGCTGGATCGAGT-----CCGAGCCCAACGCGTGTACGCTTGATCCGCATTC 966
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324

QY 967 GAACCGCAAAACGAGACAGAGCGAGCGCGCGCGCGCGCTGCGCGGTGGCGGAAGCGCC 1026
   ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIaGluGlnGlu 344

QY 1027 GCCGCGCGC 1035
   |||
Db 345 ArgGlnAla 347

RESULT 5
US-10-919-182-6
; Sequence 6, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACIC USING NITRILASE MUTANTS
; FILE REFERENCE: CI2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B2 and H9
US-10-919-182-6

Alignment Scores:
Pred. No.: 2.47e-48 Length: 369
Score: 807.00 Matches: 160
Percent Similarity: 61.8% Conservative: 52
Best Local Similarity: 46.6% Mismatches: 127
Query Match: 42.9% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-6 (1-369)

QY 16 ACGAAGTATCGCGCGCGGTGCGAGCCCGCGGTGTTCTCGATCTCGACCGCAC 75
   :::::::::: ||| ||||| ||||| ::||| ||| |||
Db 6 SerLySpHeLeuAlaIaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThr 25

QY 76 GTCGAGAAAGCGATCGGCCTGATCGAGCGCGGCCAAGCAGCAGCAGCGCTGATCGCA 135
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 26 IleAspLySerIleGlyIleIleGluGlnAlaIaGlnLySGLyAlaSerLeuIleAla 45

QY 136 TTCCCGAGACTTGAGATTCCCGGCTATCCCTTTTGATATGCTGGCGCGCGCTTG 195
   ||||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
Db 46 PheProGluValPheIleProGlyTrpProTyTrpAlaTrpLeuGlyAspValLysTyr 65

QY 196 GGCATGCGCTTCGTCAGCGCTATTTCGAGAAATTCGCTCGCGCGCGCAGCAAGCAGTGG 255
   ::||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85

QY 256 CAGGCCCTGGCGGATCGGCGCGGCCCGCCAGCATGTCGTCGCGCGCTATAGCGAG 315
   ::||| ||| ||||| ||||| ||||| ::||| ::||| ||||| ||||| ||||| |||||
Db 86 ArgArgLeuGlnLeuAlaIaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105

QY 316 CGCGCGCGCGCAGCCTTATATGGCCAGCGCATCTTCGCCCCGATGGCGATCGATC 375
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125
```

QY 376 GCCGCGCGCGCAAGCTCAAGCCTTACCCTATCGCGAGCGCACCCGTGTTCCGCGAGGAGAC 435  
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145  
QY 436 GCGACCATCTCGCGGTGCAAGATACCGCCATCGGCGCGCTCGCGCGCTGTGCTGG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrp 165  
QY 496 GAGCACATCCAGCCATTGTGAAATACGCCCATGTACCGCGCGAGCAAGACAGATCCAGCTC 555  
Db 166 GluHisPheGlnProLeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisVal 185  
QY 556 GCGTCGTGCGCGAGCTTCAAGCTCTATCGCGGATGGCCTATGCGTCCGACCGGAGTTC 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205  
QY 616 AATACCGCGCAAGCAGATCTACCGCGTCAAGGCGGCTGCTACGTCGTGCGTCTGTC 675  
Db 206 AsnAlaThrValAlaArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225  
QY 676 GCGACCGTTTCGCGCGAGATGATCAAGGTATTGGTGGATACGCGCGACAGAGATGTTTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245  
QY 736 CTCGAAGCGCGCGCGGTTTTCATGATTTTCGAGCCCGACGCGCGCGCTGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTyrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
QY 796 CCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCCGATATCGACCTCGGATGATCGCG 855  
Db 266 ProLeuAlaGlnAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGlnIleLeu 285  
QY 856 TTGGCCAAAGCGCGCGCATCCGCGCGGCGCACTATTCAAGCGCGCGCGAGTAAAGCGGCTG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValIleuSerVal 305  
QY 916 CTGCTGATCGAGT-----CCGCGCCCAACGCGTGTCAAGCTTGATCGCGCATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg--IleGlyIleAspGlyArgLeu 324  
QY 967 GAACCGCAAAAGAGGACAGGAGGCGCGCGCGCGCTGCGCGTGGTGGCGAAAGCGCC 1026  
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIleGluGlnIu 344  
QY 1027 GCCGCGCGC 1035  
Db 345 ArgGlnAla 347

RESULT 6  
US-10-919-182-12  
; Sequence 12, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cossimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 12  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change  
US-10-919-182-12  
Alignment Scores:

Pred. No.: 3.39e-48 Length: 369  
Score: 805.00 Matches: 160  
Percent Similarity: 61.5% Conservative: 51  
Best Local Similarity: 46.6% Mismatches: 128  
Query Match: 42.8% Indels: 4  
DB: 6 Gaps: 2  
US-09-751-299-1 (1-1041) x US-10-919-182-12 (1-369)  
QY 16 ACGAAGTATCGCGCGCGGTGAGCGCGCGGTGTTCTGATCTCGACCGCACA 75  
Db 6 SerLysPheLeuAlaAlaThrValGlnAlaGluProValTTrpLeuAspAlaAspAlaThr 25  
QY 76 GTCGAAGAAGCATCGGCTGTATCGACGAGCGCGCAAGCAGACGTCGCTGATCGCA 135  
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaAlaGlnLysGlyAlaSerLeuIleAla 45  
QY 136 TTCCAGAGACTTGATTCGCGGCTATCCCTTTGGATATGCTGGAGCGCGGCTTGG 195  
Db 46 PheProGluValPheIleProGlyTyrProTyrTrpAlaTrpLeuGlyAspValLysTyr 65  
QY 196 GGCATGCGCTTCGTCAGCGCTATTTCGAATTCGTCGTCGCGCGCAGCAGCAGTGG 255  
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMet 85  
QY 256 CAGGCGCTGCGCGGATGCGAGCGCGCGCGCAAGCATGATGTCGTCGCGCTATAGCGAG 315  
Db 86 ArgArgLeuGlnLeuAlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105  
QY 316 CCGCGCGCGCGCAGCCTCTATATGCGCCAGCGCATCTTCGCGCGCGATGCGATCTGATC 375  
Db 106 ArgGlnAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
QY 376 GCGCGCGCGCGCAAGCTCAAGCCTTACCCTATCGAGCGCGCGCGCGCGCGAGGAGAC 435  
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145  
QY 436 GCGACCATCTCGCGGTGCAAGATACCGCCATGTACGCGCGCGCGCGCGCTGTGCTGG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrp 165  
QY 496 GAGCACATCCAGCCATTGTGAAATACGCCCATGTACGCGCGCGCGCGCGCGCTGCGCGAG 555  
Db 166 GluHisPheGlnProLeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisVal 185  
QY 556 GCGTCGTGCGCGAGCTTCAAGCTCTATCGCGGATGGCCTATGCGTCCGACCGGAGTTC 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205  
QY 616 AATACCGCGCAAGCAGATCTACCGCGTCAAGGCGGCTGCTACGTCGTGCGTCTGTC 675  
Db 206 AsnAlaThrValCysArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225  
QY 676 GCGACCGTTTCGCGCGAGATGATCAAGGTATTGGTGGATACGCGCGAGAGATGTTTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245  
QY 736 CTCGAAGCGCGCGCGGTTTTCATGATTTTCGAGCCCGACGCGCGCGCTGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTyrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
QY 796 CCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCCGATATCGACCTCGGATGATCGCG 855  
Db 266 ProLeuAlaGlnAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGlnIleLeu 285  
QY 856 TTGGCCAAAGCGCGCGCATCCGCGCGGCGCACTATTCAAGCGCGCGCGAGTAAAGCGGCTG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValIleuSerVal 305  
QY 916 CTGCTGATCGAGT-----CCGCGCCCAACGCGTGTCAAGCTTGATCGCGCATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg--IleGlyIleAspGlyArgLeu 324

QY 967 GAACCGCAAAAGAGAGACAAGGGCGCGCCCGCTGCGCGTGTGGCGGAAGCGCC 1026  
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIaGluGlnGlu 344  
QY 1027 GCCGCCCGCG 1035  
Db 345 ArgGlnAla 347

RESULT 7

US-10-919-182-14  
; Sequence 14, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 14  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change  
US-10-919-182-14

Alignment Scores:  
Pred. No.: 3.98e-48 Length: 369  
Score: 804.00 Matches: 160  
Percent Similarity: 61.8% Conservative: 52  
Best Local Similarity: 46.6% Mismatches: 127  
Query Match: 42.7% Indels: 4  
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-14 (1-369)

QY 16 ACGAAGTATCGCGGGCGGTGACAGCGCGCGGTGTTCTCGATCTCGACCGCACA 75  
Db 6 SerLysPheLeuAlaAlaThrValGlnAlaGluProValTyrLeuAspAlaAspAlaThr 25  
QY 76 GTCGAGAAAGCGATCGCGCTGATCGAGCAGCGCGCCACAGCAGCAGTGGCGCTGATCGCA 135  
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaAlaGlnLysGlyAlaSerLeuIleAla 45  
QY 136 TTCCCAAGACTTGATTCGCCGCTATCCCTTTTGATATGGCTGGCGCGCGCTTG 195  
Db 46 PheProGlnValPheIleProGlyTyrProTyrTyrPalatPheGlyAspValLysTyr 65  
QY 196 GGCATGCGCTTGTCTCGACGCTATTTCAGAAATTGCTGTCGCGCGCAGCAGCAGTGG 255  
Db 66 SerLeuSerPheThrSerArgTyrHisGlnAsnSerLeuGluLeuGlyAspAspArgMet 85  
QY 256 CAGGCCCTGGCGGATCGCGCGCGCGCCAGCGCATGTCGTGAGCCGCTATAGCGAG 315  
Db 86 ArgArgLeuGlnLeuAlaAlaArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105  
QY 316 CGCGCGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGGCCCGCATGGCATGTATC 375  
Db 106 ArgGlnAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
QY 376 GCCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGGTTCGGCGGAGGAGAC 435  
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145  
QY 436 GGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGCGCGCTCTGTGCTGG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTyrP 165

QY 496 GAGCATCCAGCCATTGTCGAATACGCCATGTACGCCCGCCGACGACAGAGTCCACGTC 555  
Db 166 GluHisLysGlnProLeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisVal 185  
QY 556 GCGTCGTGGCCGAGCTTCAGCCTTATCGCGGATGGCTATGCGCTCGAGCCGAGGTC 615  
Db 186 AlaSerTyrProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGlnAla 205  
QY 616 AATACCGCCGCAAGCCAGATTAACGGTCGAGCGCGCGCTGCTACGTGCTGGCTGTC 675  
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225  
QY 676 GCGACCGTTTCGCCGAGATGATCAAGGTATTGTTGATACGCCCGACAGAGATGTTTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245  
QY 736 CTCAAGCGCGCGCGGTTTTCATGATTTTCGGCGCGCGCGCGCGCGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTyrPalArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
QY 796 CCGCTCCCGAGACCGAAGGAGACTGTGTCGCGCATATCGACCTCGCATGATCGCG 855  
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285  
QY 856 TTGGCCAGCGCGCGCGCGCATCCGCGCGCGCATATTCACGCGCCGACGTACCGCGCTG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305  
QY 916 CTGCTGATCGACGT-----CCGCGCCAACCGCTTCACGCTTGATCCCGCATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324  
QY 967 GAACCGCAAAAGAGAGAGAGCGCGCGCGCGCGCGCGCGCTGTCGCGCGAAGCGCC 1026  
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaAlaGluGlnGlu 344  
QY 1027 GCCGCCCGCG 1035  
Db 345 ArgGlnAla 347

RESULT 8

US-10-919-182-8  
; Sequence 8, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase B4  
US-10-919-182-8

Alignment Scores:  
Pred. No.: 1.42e-47 Length: 369  
Score: 796.00 Matches: 160  
Percent Similarity: 60.9% Conservative: 49  
Best Local Similarity: 46.6% Mismatches: 130  
Query Match: 42.3% Indels: 4  
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-8 (1-369)

```
QY 16 ACGAAGTATCGCGCGCGGTGTCAGCGCGCGGTGTCTCTGATCTGACCGCACA 75
Db 6 SerLysPheLeuAlaIaThrValGlnAlaGluProValTyrLeuAspAlaAspAlaThr 25
QY 76 GTCGAGAAAGCGATCGGCTTGATCGAGCAGCGCGCAAGCAGAGACGTGCGCTGATCGCA 135
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaIaGlnLysGlyAlaSerLeuIleAla 45
QY 136 TTCCAGAGACTTGATTCCTCCGGCTATCCCTTTTGATATGCTGGCGCGCGCTTGG 195
Db 46 PheProGluValPheIleProGlyTyrProTyrTyrPalaTyrPheGlyAspValLysCys 65
QY 196 GGCATGCGCTTCGTCAGCGCTATTTCGAGATTGCTGCGCGCGCAGCAAGCAGTGG 255
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspArgMet 85
QY 256 CAGGCCCTGGCGGATCGGCGCGCGCGCGCGCGCATGTCATGTCGTGCGCGCTATAGCGAG 315
Db 86 ArgArgLeuGlnLeuAlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105
QY 316 CGCGCGCGCGCGCTCTATATGCGCAGCGCATCTTCGGCCCGATGCGCATCTGATC 375
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125
QY 376 GCGCGCGCGCGCAGCTCAAGCCTACCCATCGGAGCGCAGCAGCGCTGTCGGCGAGGAGAC 435
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsn 145
QY 436 GGCAGCCATCTCGCGGTGCAGCATACCGCCATCGGCGCGCTCGCGCGCTGTGCTGG 495
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysThr 165
QY 496 GAGCACATCCAGCCTTGTGAAATACGCCCATGTACCGCGCGCGCTGCTACGTGCGCTGTC 555
Db 166 GluHisPheGlnProLeuSerLysIleMetMetTyrSerLeuGlyGluGlnValHisVal 185
QY 556 GCGTGTGCGCGAGCTTCAGCCTCTATCGCGCGCATGCGCTATGCGCTCGACCGGAGTGC 615
Db 186 AlaSerTyrProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205
QY 616 AATACCGCGCGCAGCAGATCTACGCGGTGCGAGCGCGCTGCTACGTGCGCTGTC 675
Db 206 AsnAlaThrValIleArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225
QY 676 GCGACCGTTTCGCGCGAGATGATCAAGGTATTGCTGATACGCGCGCAGAGATGTTTC 735
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245
QY 736 CTCAAGGCGCGCGGTGTTTGCATGATTTTTCGAGCCCGACGCGCGCGCTGCGCGAG 795
Db 246 LeuProGlnGlyCysGlyTyrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265
QY 796 CCGCTCCCGAGAGACGAGAGGAGTCTGTGTCGCCGATATCGACCTCGGATGATCGCG 855
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285
QY 856 TTGGCCAGGCGCGCGCGATCCGCGCGCGCTATTCACGGCGCGCAGCTAACCGGCTG 915
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValIleuSerVal 305
QY 916 CTGCTGATCGACGT-----CCGCGCCCAACGCGTGTGTCACGCTTGATGCGCATTC 966
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324
QY 967 GAACCGCAAAACGAGACAAGGGCGAGCGCGCGCTGCGCGTGTGCGGAAAGCGCC 1026
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIaGluGlnGlu 344
QY 1027 GCCGCGCGC 1035
Db 345 ArgGlnAla 347
```

RESULT 9

```
US-11-096-568A-20687
; Sequence 20687, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20687
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(333)
; OTHER INFORMATION: Ceres Seq. ID no. 12386686
US-11-096-568A-20687

Alignment Scores:
Pred. No.: 2.23e-22 Length: 333
Score: 430.50 Matches: 117
Percent Similarity: 50.6% Conservative: 52
Best Local Similarity: 35.0% Mismatches: 128
Query Match: 22.9% Indels: 37
DB: 7 Gaps: 8

US-09-751-299-1 (1-1041) x US-11-096-568A-20687 (1-333)
QY 16 ACGAAGTATCGCGCGCGGTGTCAGCGCGCGGTGTCTCTGATCTGACCGCACA 75
Db 10 ThrThrAlaArgValThrValValGlnAlaSerSerValPheTyrAspThrProAlaThr 29
QY 76 GTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAGACGTGCGCTGATCGCA 135
Db 30 LeuAspLysAlaGluLysLeuValAlaGluAlaIaGlyTyrGlySerGlnLeuValLeu 49
QY 136 TTCCAGAGACTTGATTCCTCCGGCTATCC-----TTTGGATATGCGCTGGC 183
Db 50 PheProGluValPheValGlyGlyTyrProHisGlySerThrPheGlyLeuValIaGly 69
QY 184 ---GCCGCGCTTGGGCGATGCGCTGCTGCCAGCGCTATTTCGAGAAATTCGCTGCGC 240
Db 70 AsnArgThrAlaLysGlyLysGluAspPheGlnLysTyrHisAlaSerAlaIleAspVal 89
QY 241 GGCAGCAGCAGTGGCAGGCGCGCTGCGCGATGCGCGCGCGCGCGCATGCTGCTG 300
Db 90 ProGlyProGluValSerArgLeuSerAlaLeuAlaGlyLysTyrLysValPheLeuVal 109
QY 301 GCCGCTATAGCGAGCGCGCGCGCGCGCTCTATATGCGCAGCGCATCTTCGGCCCC 360
Db 110 IleGlyValValGluArgAlaGlyTyrThrLeuTyrAsnThrValLeuSerPheAspPro 129
QY 361 GATGCGATGTATCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGAGCGCACCGTG 420
Db 130 LeuGlyLysTyrLeuGlyLysHisArgLysValMetProThrAlaLeuGluArgValPhe 149
QY 421 TTCGCGAGGAGAGCGCAGCAGCATCTCGCGTGCACGATACCGCGCATCGGCGCTGCG 480
Db 150 TyrGlyPheGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysMetGly 169
QY 481 GCGCTGTGCTGGAGCAGCATTCAGCATTTGTGGAATACGCCATGTACGCCGCGAC 540
Db 170 AlaLeuIleCysTyrPgluAsnArgMetProLeuLeuArgThrAlaMetTyrAlaLysGly 189
QY 541 GAACAGGTCCAGCTGCGCTGTGCGCGAGCTTCAGCCTCTATCGCGGATGCGCTATGCG 600
Db 190 IleGluIleTyrCysAla----- 195
QY 601 CTCGAGCGAGGTCAATACCGCGCGCAGCCAGATC-----TACGCGGTC 645
```



Db 196 -----ProthrValaAspCysmetProthrTrpleuserSermetThrHisIleAlaLeu 213  
QY 646 GAGGCGGCTGCTACGTGCTGCGCTGCGCG-----ACCGTT 684  
Db 214 GluGlyGlyCysPheValleuserAlaCysGlnPheCysArgArgLysAsnTyrProPro 233  
QY 685 TCGCCGAGATGATCAAGTA--TTGGTGATACGCCCGACAGAAGATGTTCTCAAG 741  
Db 234 ProProGluTyrThrPheCysGlyLeuGluGluProserProGluSerValValCys 253  
QY 742 GCCGCGCGGCTTTTGCCATGATTTTCGGGCCGACGCGCGCCCTGCGCCGAGCCGCTC 801  
Db 254 SerGlyGly-----SerValIleIleSerProleuGlyThrValLeuAlaGlyProAsn 271  
QY 802 CCGAGACCGAAGAGGACTGTGTGCGCCATATCGACCTCGGCATGATCGCGTTGGCC 861  
Db 272 TyrGluSer---GluAlaLeuLeuThrAlaAspLeuAspLeuGlyGluIleValArgAla 290  
QY 862 AAGCGCGCGCGCATCCGGCGGCCACTATTTCACGCCCGACGTACCGCGGCTGCTGCTG 921  
Db 291 LysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSerLeuValVal 310  
QY 922 GATCGACGTCGCGCCCAACGCGTCGTACGCTGTATGCCGCA 963  
Db 311 LysSerAspProLysProAlaValSerPheIleSerAlaAla 324

## RESULT 10

US-11-096-568A-20686  
; Sequence 20686, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096, 568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20686  
; LENGTH: 351  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(351)  
; OTHER INFORMATION: Ceres Seq. ID no. 12386685  
US-11-096-568A-20686

## Alignment Scores:

Pred. No.:	2.24e-22	Length:	351
Score:	430.50	Matches:	117
Percent Similarity:	50.6%	Conservative:	52
Best Local Similarity:	35.0%	Mismatches:	128
Query Match:	22.9%	Indels:	37
DB:	7	Gaps:	8

US-09-751-299-1 (1-1041) x US-11-096-568A-20686 (1-351)

QY 16 ACGAAGTATCGCGCGCGCGGTGACGCCGCGCGGTTCCTCGATCTGCACCGCACA 75  
Db 28 ThrThrAlaArgValThrValValGlnAlaSerSerValPheTyrAspThrProAlaThr 47  
QY 76 GTCGAGAAAGCGATCGCGCTGATCGACGAGCGGCCAAGCAGACGCGCTGATCGCA 135  
Db 48 LeuAspLysAlaGluLysLeuValAlaGluAlaAlaGlyTyrGlySerGlnLeuValLeu 67  
QY 136 TTCCGACAGACTTGATTCGCCGCTATCCC-----TTTGGATATGCGTGGGC 183  
Db 68 PheProGluValPheValGlyGlyTyrProHisGlySerThrPheGlyLeuValValGly 87  
QY 184 ---GCGCGGCTTGGGCGATGCGCTTCGTCACGCGTATTTCGAGAAATTCGCTGCGC 240  
Db 88 AsnArgThrAlaLysGlyLysGluAspPheGlnLysTyrHisAlaSerAlaIleAspVal 107

QY 241 GGCAGCAAGCAGTGGCAGGCGCCTTGCGGATGCGGCCCGCCGCAAGGCATGATGCTG 300  
Db 108 ProGlyProGluValSerArgLeuSerAlaLeuAlaGlyLysTyrLysValPheLeuVal 127  
QY 301 GCCGCTATAGCAGCGCGCGCGGCGCAGCCTCTATATAGGCCAGCGCATCTTGGCCCC 360  
Db 128 IleGlyValValGluArgAlaGlyTyrThrLeuTyrAsnThrValLeuSerPheAspPro 147  
QY 361 GATGGCATCTGATCGCCGCGCGCGCAAGCTCAAGCCTAACCCATGCGGAGCGCACCGTG 420  
Db 148 LeuGlyLysTyrLeuGlyLysHisArgLysValMetProThrAlaLeuGluArgValPhe 167  
QY 421 TTCGCGAGGAGAGCAGCGCAGCATCTCGCGGTGACAGATACCGCATCGGCGCCTGCGC 480  
Db 168 TrpGlyPheGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysMetGly 187  
QY 481 GCGCTCTGTTGCTGGAGCACATCCAGCCATTGTGGAATAACGCCATGTACGCCGCCGAC 540  
Db 188 AlaLeuIleCysTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyrAlaLysGly 207  
QY 541 GAACAGTCCACGTCGCGCTGTCGCCGAGCTTCAGCCTTATCGCGGCATGCGCTATGCG 600  
Db 208 IleGluIleTyrCysAla----- 213  
QY 601 CTCGACCGGAGGTCAATATCCGCCGCAAGCCAGATC-----TACCGCGTC 645  
Db 214 -----ProThrValaAspCysmetProthrTrpleuserSermetThrHisIleAlaLeu 231  
QY 646 GAGGCGGCTGCTACGTGCTGCGCTGCTGCGCG-----ACCGTT 684  
Db 232 GluGlyGlyCysPheValleuserAlaCysGlnPheCysArgArgLysAsnTyrProPro 251  
QY 685 TCGCCGAGATGATCAAGTA--TTGGTGATACGCCCGACAGAAGATGTTCTCAAG 741  
Db 252 ProProGluTyrThrPheCysGlyLeuGluGluProserProGluSerValValCys 271  
QY 742 GCCGCGCGGCTTTTGCCATGATTTTCGGGCCGACGCGCGCCCTGCGCGCGCTC 801  
Db 272 SerGlyGly-----SerValIleIleSerProleuGlyThrValLeuAlaGlyProAsn 289  
QY 802 CCGGACCGAAGAGGACTGTGTGCGCCGATATCGACCTCGGCATGATGCGCTGGCC 861  
Db 290 TyrGluSer---GluAlaLeuLeuThrAlaAspLeuAspLeuGlyGluIleValArgAla 308  
QY 862 AAGCGCGCGCGCATCCGGCGGCCACTATTTCACGCGCGCCGACGTAACGCGCTGCTG 921  
Db 309 LysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSerLeuValVal 328  
QY 922 GATCGACGTCGCGCCCAACGCGTCGTACGCTGTATGCCGCA 963  
Db 329 LysSerAspProLysProAlaValSerPheIleSerAlaAla 342

## RESULT 11

US-11-096-568A-6955  
; Sequence 6955, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096, 568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6955  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(330)  
; OTHER INFORMATION: Ceres Seq. ID no. 15169318

US-11-096-568A-6955

Alignment Scores:

Pred. No.:	3.05e-20	Length:	330
Score:	399.50	Matches:	112
Percent Similarity:	50.3%	Conservative:	57
Best Local Similarity:	33.3%	Mismatches:	136
Query Match:	21.2%	Indels:	31
DB:	7	Gaps:	11

US-09-751-299-1 (1-1041) x US-11-096-568A-6955 (1-330)

```
QY 25 CGCGGCGGCGGTGCAGCGCGCGCGGTTCCTCGATCTCGACCGCACATCGAGAA 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11 ArgAlaThrValValGlnAlaSerThrIlePheTyraSPThrProAlaThrLeuAspLys 30

QY 85 GCGATCGCGCTGATCGAGCAGCGCGCAAGCAGACGTCGCTGATCGCATTCAGAG 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 AlaGluArgLeuLeuAlaGluAlaThrSerTyrgLysSerGlnLeuValPheProGlu 50

QY 145 ACTTGATTCGCGCTATCCC-----TTTGGATATGCTGGG---GCGCG 189
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 AlaPheValGlyGlyTyrrProaRgLySerAlaPheGlyLeuSerIleGlyAsnArgThr 70

QY 190 GCTTGGGCGATCGCGCTTCCTCCAGCGCTATTTCGAATTCGCTCGTCGCGCGCAAG 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 VallysGlyArgGluGluPheArgLysTyrrHisSerAlaAlaIleAspValProGlyPro 90

QY 250 CAGTGGCAGGCGCTGGCGGATGCGCGCGCGCGCGCATGTCATGTCGTGGCGGCTAT 309
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 GluValAspArgLeuAlaAlaMetAlaGlyTyrrLysValHisLeuValMetGlyVal 110

QY 310 AGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCATCTTCGCGCGCGAGTGGCAT 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 IleGluArgAspGlyTyrrThrLeuTyrrCysThrValLeuPhePheAspSerGlnGlyHis 130

QY 370 CTGATCGCGCGCGCGCGAGCTCAAGCTTACCGCATGCGGAGCGCGCATGTTTCGCGAG 429
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 TyrLeuGlyLysHisArgLysIleMetProThrAlaLeuGluArgValIleTrpGlyPhe 150

QY 430 GGAGACGGCAGCCATCTCCGCGGTGCAGCATACCGCATCGCGCGCTCGCGCGCTGT 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 GlyAspGlySerThrIleProValPheGluThrProValGlyLysIleGlyAlaAlaIle 170

QY 490 TGCTGGAGCACATCCAGCATTTGTGAATATCCCATGTACGCGCGCGCAACAGGTC 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 CysTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyrrAlaLysGlyValGluIle 190

QY 550 CACGTGCGGTGCGCGCGAGCTTCAGCCTTATCGCGCGCATGCGCTATGCGTGGACCG 609
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 TyrCysAla-----ProThr-----AlaAspAlaArg 199

QY 610 GAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGTGCTACGTGGCG 669
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 AspValTrpGlnAlaSerMetThrHisIleAlaLeuGluGlyCysPheValLeuSer 219

QY 670 TCG-----TGGCG-----ACCGTTTCGCGCGAGATGATC-----AAG 702
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 AlaAsnGlnPheCysArgArgArgAspTyrrProProProGluTyrrValPheAlaGly 239

QY 703 GTATTGTGAT--ACGCGCGCAAGAGATGTTCTCAAGCGCGCGCGGTTTGGC 759
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 ThrGluValAspLeuThrProAspSer-----ValValCysAlaGlyGly-----Ser 255

QY 760 ATGATTTTCGGCGCGCGCGCGCGCGCTGCGCGCGCGCTCCGAGACCGAAGAGGA 819
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ValIleIleSerProLeuGlyAlaValLeuAlaGlyProAsnTyrrAsp---GlyGluAla 274

QY 820 CTGCTGTCGCGCGATATGACCTTCGCGATGTCGCTTGGCCAAGCGCGCGCGCATCCG 879
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 275 LeuIleSerAlaAspLeuAspLeuGlyLysIleAlaArgAlaLysPheAspPheAspVal 294

QY 880 GCGGCGCACTATTCAGGCGCGCGAGCGTAACGCGGCTGCTGGATGACGTCGCGGCCAA 939
```

```
Db 295 ValGlyHisTyrrSerArgProGluValLeuSerLeuThrValLysAspHisProThraSn 314
QY 940 CGCGTCGTACGCTTGATCGCGCATTCGAACCGCAAAACGAGGACAG 987
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 ProValThrPheThrSerThrSerThrLysIleGluAspLysThrLys 330
```

RESULT 12

```
US-11-096-568A-6954
; Sequence 6954, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6954
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(350)
; OTHER INFORMATION: Ceres Seq. ID no. 15169317
US-11-096-568A-6954
```

Alignment Scores:

Pred. No.:	3.07e-20	Length:	350
Score:	399.50	Matches:	112
Percent Similarity:	50.3%	Conservative:	57
Best Local Similarity:	33.3%	Mismatches:	136
Query Match:	21.2%	Indels:	31
DB:	7	Gaps:	11

US-09-751-299-1 (1-1041) x US-11-096-568A-6954 (1-350)

```
QY 25 CGCGGCGGCGGTGCAGCGCGCGCGGTTCCTCGATCTCGACCGCACATCGAGAA 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 ArgAlaThrValValGlnAlaSerThrIlePheTyraSPThrProAlaThrLeuAspLys 50

QY 85 GCGATCGCGCTGATCGAGCAGCGCGCGCAAGCAGACGTCGCTGATCGCATTCAGAG 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 AlaGluArgLeuLeuAlaGluAlaThrSerTyrgLysSerGlnLeuValPheProGlu 70

QY 145 ACTTGATTCGCGGCTATCCC-----TTTGGATATGCTGGG---GCGCG 189
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 71 AlaPheValGlyGlyTyrrProaRgLySerAlaPheGlyLeuSerIleGlyAsnArgThr 90

QY 190 GCTTGGGCGATCGCGCTTCGTCAGCGCTATTTCGAATTCGCTCGCGCGCGCGAG 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 VallysGlyArgGluGluPheArgLysTyrrHisSerAlaAlaIleAspValProGlyPro 110

QY 250 CAGTGGCAGGCGCTGGCGGATGCGCGCGCGCGCGCATGTCATGTCGTGGCGGCTAT 309
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 GluValAspArgLeuAlaAlaMetAlaGlyTyrrLysValHisLeuValMetGlyVal 130

QY 310 AGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCATCTTCGCGCGCGAGTGGCAT 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 IleGluArgAspGlyTyrrThrLeuTyrrCysThrValLeuPhePheAspSerGlnGlyHis 150

QY 370 CTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCGCATGCGGAGCGCACCGTTCGCGAG 429
    : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 TyrLeuGlyLysHisArgLysIleMetProThrAlaLeuGluArgValIleTrpGlyPhe 170

QY 430 GGAGACGGCAGCCATCTCCGCGGTGCAGCATACCGCATCGCGCGCGCTCGCGCTGT 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 GlyAspGlySerThrIleProValPheGluThrProValGlyLysIleGlyAlaAlaIle 190

QY 490 TGCTGGAGCACATCCAGCATTTGTGAATATCGCCATGTACGCGCGCGCAACAGGTC 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 191 CysTirpGluaSnaArgMetProLeuLeuArgThrAlaMetTyrAlaIalysGlyValGluIle 210  
QY 550 CACGTCGCGTCGTGGCCGAGCTTCAGCCTCTATCGCGCAGTGGCCTATGCGCTCGACCG 609  
Db 211 TyrCysAla-----ProThr-----AlaAspAlaArg 219  
QY 610 GAGGTCAATACCGCCGCAAGCCAGATCTACGCGTCGAGGCGCGCTCTACGTGTCGCG 669  
Db 220 AspValTirpGluaIaSerMetThrHisIleAlaLeuGluGlyCysPheValLeuSer 239  
QY 670 TCG-----TGGCGG-----ACCGTTTCGCGGAGATGATC-----AAG 702  
Db 240 AlaAsnGlnPheCysArgArgArgAspTyrProProProGluTyrValPheAlaGly 259  
QY 703 GTATTGTGGAT---ACGCCCGACAAGAGATGTTCTCAAGCGCGCGCGGTTTGGCC 759  
Db 260 ThrGluValAspLeuThrProAspSer-----ValValCysAlaGlyGly-----Ser 275  
QY 760 ATGATTTTCGGGCGCCGACGCGCGCGCTGCGCGAGCCGCTCCCGAGACCGAAGAGGA 819  
Db 276 ValIleIleSerProLeuGlyAlaValLeuAlaGlyProAsnTyrAsp---GlyGluAla 294  
QY 820 CTGCTGTCGCGCGATATCGACCTCGGCATGATCGCGTGGCCCAAGCGCGCCGATCCG 879  
Db 295 LeuIleSerAlaAspLeuAspLeuGlyGluIleAlaArgAlaIalysPheAspPheAspVal 314  
QY 880 GCGGGCCACTATTACGCGCGCCGACGTACGCGCGCTGCTGATCGACGTCCGCGCCAA 939  
Db 315 ValGlyHisTyrSerArgProGluValLeuSerIleThrValIlysAspHisProThrAsn 334  
QY 940 CGCGTCGTACGCTTGATGCCCATTCGAACCGCAAAACGAGACAAG 987  
Db 335 ProValThrPheThrSerThrSerThrLysIleGluAspLysThrLys 350

RESULT 13  
US-11-096-568A-6956  
; Sequence 6956, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6956  
; LENGTH: 233  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(233)  
; OTHER INFORMATION: Ceres Seq. ID no. 15169319  
US-11-096-568A-6956

Alignment Scores:  
Pred. No.: 1.99e-14 Length: 233  
Score: 315.00 Matches: 88  
Percent Similarity: 51.6% Conservative: 40  
Best Local Similarity: 35.5% Mismatches: 94  
Query Match: 16.7% Indels: 26  
DB: 7 Gaps: 9

US-09-751-299-1 (1-1041) x US-11-096-568A-6956 (1-233)  
QY 274 GCCCGCCGCAAGCATGATGTCGTGGCCGGCTATAGCAGCGCGCGGCGCAGCCTC 333  
Db 2 AlaGlyLysTyrLysValHisLeuValMetGlyValIleGluArgAspGlyTyrThrLeu 21  
QY 334 TATATGGCCAGCGCATCTTCGCGCCCGATGGCGATCTGATCGCGCGCGCGCAAGCTC 393  
Db 22 TyrCysThrValLeuPhePheAspSerGlnGlyHisTyrLeuGlyLysHisArgLysIle 41

QY 394 AAGCTTACCCATGCGGAGCGCACCCGTGTTCCGCGAGGAGACCGGACCATCTCGCGTG 453  
Db 42 MetProThrAlaLeuGluArgValIleTirpGlyPheGlyAspGlySerThrIleProVal 61  
QY 454 CACGATACCGCCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAGACATCCAGCATTG 513  
Db 62 PheGluThrProValGlyLysIleGlyAlaAlaIleCysTirpGluaSnaArgMetProLeu 81  
QY 514 TCGAAATACGCATGTACGCCGCGCAGCAAGATCCACGTCCGCTGTCGCGAGCTTC 573  
Db 82 LeuArgThrAlaMetTyrAlaLysGlyValGluIleTyrCysAla-----ProThr--- 98  
QY 574 AGCCTCTATCGCGCATGCGCTTATGCGCTCGAACCGAGTCAATACCGCGCAAGCCAG 633  
Db 99 -----AlaAspAlaArgAspValTirpGluaIaSerMetThr 110  
QY 634 ATCTACGCGGTGAGGCGCGCTGCTACGTGTCGCGTGC-----TGGCGG----- 678  
Db 111 HisIleAlaLeuGluGlyGlyCysPheValLeuSerAlaAsnGlnPheCysArgArgArg 130  
QY 679 -----ACCGTTTCGCGGAGATGATC-----AAGTATTGTGGAT---ACGCCGAC 723  
Db 131 AspTyrProProProProGluTyrValPheAlaGlyThrGluValAspLeuThrProAsp 150  
QY 724 AAGGATGTTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGCGCGCGCGCGCGC 783  
Db 151 Ser-----ValValCysAlaGlyGly-----SerValIleIleSerProLeuGlyAla 166  
QY 784 GCCCTGCGCGAGCGCGCTCCGCGAGACCGAAGAGGAGTCTGTGCGCGCATTCGACCTC 843  
Db 167 ValLeuAlaGlyProAsnTyrAsp---GlyGluAlaLeuIleSerAlaAspLeuAspLeu 185  
QY 844 GGCATGATCGCGTGGCCAGGCGCGCGCGATCCGCGCGCCACTATTACGCGCGCGAC 903  
Db 186 GlyGluIleAlaArgAlaLysPheAspPheAspValValGlyHisTyrSerArgProGlu 205  
QY 904 GTAACGCGGCTGCTGTCGATCGACGTCCGCGCCCAACGCGTTCACGCTTGATCGCGCA 963  
Db 206 ValLeuSerLeuThrValIlysAspHisProThrAsnProValThrPheThrSerThrSer 225  
QY 964 TTGACCGCGCAAAACGAGACAAG 987  
Db 226 ThrLysIleGluAspLysThrLys 233

RESULT 14  
US-11-096-568A-20688  
; Sequence 20688, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Therby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20688  
; LENGTH: 193  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(193)  
; OTHER INFORMATION: Ceres Seq. ID no. 12386687  
US-11-096-568A-20688

Alignment Scores:  
Pred. No.: 2.7e-12 Length: 193  
Score: 284.00 Matches: 75  
Percent Similarity: 51.0% Conservative: 28  
Best Local Similarity: 37.1% Mismatches: 67  
Query Match: 15.1% Indels: 32

DB: 7 Gaps: 6  
US-09-751-299-1 (1-1041) x US-11-096-568A-20688 (1-193)  
QY 397 CCTACCCATGCGGAGCGCAACCGTGTTCGGCGAGGAGACCGCAGCCATCTCCGGGTGCAC 456  
Db 2 ProThrAlaLeuGluArgValPheTrpGlyPheGlyAspGlySerThrIleProValTyr 21  
QY 457 GATACCGCCATCGGCGCGCTTCGGCGCTGTGTTGCTGGAGACACATCCAGCCATTGTCG 516  
Db 22 AspThrProIleGlyLysMetGlyAlaLeuIleCysTrpGluAsnArgMetProLeuLeu 41  
QY 517 AAATACGCCATGTACGCGCGCGACGAACAGTCCACGTCGCTGCGCGCAGCTTCAGC 576  
Db 42 ArgThrAlaMetTyrAlaLysGlyIleGluIleTyrCysAla----- 55  
QY 577 CTCTATCGCGCATGCGCTATGCGCTCGACCGGAGGTCAATACCGCGCAAGCAGATC 636  
Db 56 -----ProThrValAspCysMetProThrTrpLeu 65  
QY 637 -----TACGCGTTCGAGGCGCGCTGCTACGTGCTGCGCTGCGCG--- 678  
Db 66 SerSerMetThrIleAlaLeuGluGlyCysPheValLeuSerAlaCysGlnPhe 85  
QY 679 -----ACCGTTTCGCGCGAGATGATCAAGTA--TTGTTGATACG 717  
Db 86 CysArgArgLysAsnTyrProProProProGluTyrThrPheCysGlyLeuGluGlu 105  
QY 718 CCGGACAAGAGATGTTCTCAAGCGCGCGCGGTTTTCATGATTTTCGGCGCGAC 777  
Db 106 ProSerProGluSerValValCysSerGly-----SerValIleIleSerProLeu 123  
QY 778 GCGCGCGCTGCGCGAGCGCTCCCGAGACCGAAGGAGACTGTGTCGCGCATATC 837  
Db 124 GlyThrValLeuAlaGlyProAsnTyrGluSer---GluAlaLeuLeuThrAlaAspLeu 142  
QY 838 GACCTCGCATGATCGCGTGGCCCAAGCGCGCGCGCATCCGGCGCGCATATCAGCG 897  
Db 143 AspleuGlyGluIleValArgAlaLysPheAspPheAspValValGlyHisTyrSerArg 162  
QY 898 CCGGACGTAAACGCGCTGCTGCTGATCGACGTCCGCGCAACGCGTCTACCGCTGAT 957  
Db 163 ProGluValLeuSerLeuValValLysSerAspProLysProAlaValSerPheIleSer 182  
QY 958 GCCGCA 963  
Db 183 AlaAla 184  
RESULT 15  
US-11-096-568A-20255  
; Sequence 20255, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20255  
; LENGTH: 449  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURES:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(449)  
; OTHER INFORMATION: Ceres Seq. ID no. 12381092  
US-11-096-568A-20255  
Alignment Scores: 2.65e-09 Length: 449  
Pred. No.: 241.00 Matches: 117

Percent Similarity: 36.4% Conservative: 23  
Best Local Similarity: 30.4% Mismatches: 138  
Query Match: 12.8% Indels: 107  
DB: 7 Gaps: 20  
US-09-751-299-1 (1-1041) x US-11-096-568A-20255 (1-449)  
QY 1037 TGCGCGCGCGCGCTTTCGCCACACCGCGCAGCGCGCGCTGCTGCTCTCG 978  
Db 89 CysAlaProGlyAlaAla-ProProAlaAlaProPheProAlaArgHisProAlaArgHisAr 108  
QY 977 TTTTCGGTTGATGCGGCATCAAGCGTGAAGCGCGTGGCGCGGACGTGCATCCAGC 918  
Db 108 gArgAspArgArgLeuArgHisGlyProLeuArgArgArgGlyProGlnLeuAlaAla 128  
QY 917 AGCAGC----- 911  
Db 128 alaAlaAspGlyGluAspValLeuArgLeuArgHisThrValLeuArgGlyAlaGly 148  
QY 910 -----GCGTTACGTGCGCGCGCGTGAATAGCGCGCGCGA 876  
Db 148 yAlaGlyAlaArgValProAlaGlnSerArgArgValProGlyProGlnAlaGlyGly 168  
QY 875 TCGCGCGCGCTTGC----- 860  
Db 168 uArgProHisProGlyGlnArgProHisAlaAlaArgArgLeuArgProLeuHisAlaAla 188  
QY 859 -CCAACGCGATCATGCCAGGTGCATATCGCGCACAGCATCCCT-----CTTCG 810  
Db 188 aProGlyArgSerArgArgSerGlnThrArgArgProProValProSerArgArgLeuAr 208  
QY 809 GTCTCCGGGAGCGGCTCGGCCAGGCGCGCGCTCGGCCGAAATCATGCAAAACCG 750  
Db 208 gSerPro-----ArgArgProArgAlaAlaAlaGluGlnGlySerAr 222  
QY 749 CCGCGCGCTTGAGAACATCTCTGTTCGCGGTAT-----CCACCAATACCTTGATC 696  
Db 222 gArgAlaPro-----ProCysAlaSerHisSerAlaProProAlaAlaLeuAr 238  
QY 695 ATCT--CCGCGAAACGCTCGCGCACAGCGCACGATAGCAGCGCGCTCGACCGCG 639  
Db 238 gSerLeuProArgThrArgArgArgCysProProArg--GlyArgProArgArgAr 257  
QY 638 TAGATCTGCGCTTGGCGGTATTTGACCTCCGCTCCGAGCGCATAGGCCATGCCGATAG 579  
Db 257 gArgArgProArgArgArgProArgProProArgArgArgArgProAla-----Ar 275  
QY 578 AGCTGAAGCTCGGCCACAGCGCGACGTGACTTTCGTGCGCGCGCTACATGGCGTAT 519  
Db 275 gGlyLeuProArgSerArgThrArgSerTrp-----GlyArgArgThrTrpArgAr 293  
QY 518 TTCGACATGCTGATGTGCTCCAGCACAGACGCGCGCGCGCGATGCGCGTA 459  
Db 293 gArgSerSerProGlyAlaGlyThrThrSerProTrpThrGlyGlyAla--TrpAlaTr 312  
QY 458 TCGTGACCGCGAGATGGCTGCGCTTCCTCGCCGAACAGTGCCTCGCATGGTA 399  
Db 312 p-----Cys-----SerThrArgCys----- 317  
QY 398 GGCTTAGCTTGGCGCGCGCGCATCATCGGCAGCGCGCATCGCCTGGCCC 339  
Db 318 -----SerThrAlaAlaArgArgSerGly-----GlyArgThrAlaArgArgPr 332  
QY 338 ATATAGAGGCTGCCCGCGCGCGCTGC---TATAGCGGCCACGACATGCATGCCGTGG 282  
Db 332 oSerThrAlaSerSerProSerSerArgSerSerAlaSerArgArgArgCysAlaTh 352  
QY 281 CGGCGGCGCGCATCCGCCAGGCGCTGCCACTGTGCTGCGCGCGCACGCAATTCTCG 222  
Db 352 rSerSerPro-----AlaSerSerArgArgThrProAr 363  
QY 221 AAATAGCGCTGACGAAGCGCATGCCCAAGCGCGCGCGCGCATATCCAAAGGGA 162  
| : : : : : | : : : : : |



Db 363 gGlyAlaSerAlaArgAlaAla-----SerArgProThrProSerSerAlaAl 379  
Qy 161 TAGCCGGGAATCCAAAGTCTCTGGGAATGCGATCAGCGCACGTCCTGCTTGCCCGCCTGC 102  
Db 379 aSerThrGlyThrHisProAlaGly-----GlyAlaProAlaValHisProGlyAl 396  
Qy 101 TCGATCAGCGCGATCGCTTCTCGACTGTGCCGTCGAGATCGAGAACACCGCGCGGCC 42  
Db 396 aAlaAlaGlyArg-----GlyArgArgArgGlyAlaArgArgGly 410  
Qy 41 TGCACCGCGCGCGC 29  
Db 410 uGlyArgProArg 414

Search completed: April 27, 2006, 01:53:07  
Job time : 29.2628 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:29:49 ; Search time 4902.58 Seconds  
(without alignments)  
9934.641 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgccgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_est3: \*  
4: gb\_est4: \*  
5: gb\_est5: \*  
6: gb\_est6: \*  
7: gb\_est7: \*  
8: gb\_est8: \*  
9: gb\_est9: \*  
10: gb\_est10: \*  
11: gb\_est11: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	155	14.9	888	9	AQ329770 nbxb0045P
2	153	14.7	622	9	AQ329769 nbxb0045P
3	140.8	13.5	699	8	DR631467 EST102159
4	140.8	13.5	732	8	DR631867 EST102199
5	140.8	13.5	810	8	DR629486 EST101961
6	140.8	13.5	841	8	DR632130 EST102225
7	140.8	13.5	856	8	DR622046 EST101217
8	140.8	13.5	858	8	DR630522 EST102065
9	129.8	12.5	655	6	CF453420 fct536.1.1
10	129.4	12.4	840	7	CN808140 Blood_EST
11	129.4	12.4	998	7	CN808249 Blood_EST
12	127.6	12.3	769	8	DR609823 EST99951
13	122.6	11.8	840	8	DR709497 Asn_11063
14	116.6	11.2	580	9	AQ160684 mgxb0006L
15	109.4	10.5	804	8	DR709485 Asn_11049
16	108	10.4	772	6	CD459513 Fg09_03j1
17	101.2	9.7	551	7	CO149108 EST824161
18	101.2	9.7	553	7	CO149186 EST824239
19	99.6	9.6	610	7	CO134433 EST829104
20	97.8	9.4	546	7	CO139118 EST833789
21	91	8.7	536	8	DR622501 EST101262
22	91	8.7	561	8	DR622709 EST101283

23	91	8.7	595	8	DR626172	DR626172	EST101630
24	91	8.7	633	8	DR622498	DR622498	EST101262
25	91	8.7	653	8	DR626120	DR626120	EST101624
26	91	8.7	758	8	DR625821	DR625821	EST101594
27	90.6	8.7	515	7	CO137063	CO137063	EST831734
28	89.2	8.6	409	1	AW711275	AW711275	fla1ne.f
29	88.6	8.5	866	8	DR621807	DR621807	EST101193
30	88.2	8.5	434	1	AF408416	AF408416	AF408416
31	83.8	8.0	937	8	DR630423	DR630423	EST101475
32	83.6	8.0	903	8	DR630423	DR630423	EST102055
33	81	7.8	532	2	BG278308	BG278308	a3d02np.r
34	80.8	7.8	838	10	CG811199	CG811199	FSAAN59TR
35	80.2	7.7	418	1	AW713376	AW713376	g8c04ne.f
36	75.4	7.2	417	1	AW710819	AW710819	e6h01ne.f
37	75.4	7.2	672	8	DR629554	DR629554	EST101968
38	74.2	7.1	441	10	CW493581	CW493581	fbab001f2
39	74.2	7.1	546	10	CW149572	CW149572	104_548.1
40	72.6	7.0	383	8	DR625648	DR625648	EST101577
41	72.4	7.0	419	2	BF072824	BF072824	NCSP7C3T3
42	71.4	6.9	412	1	AI329382	AI329382	b4a10ne.f
43	70.6	6.8	359	1	AW710710	AW710710	e5f03ne.f
44	68.8	6.6	744	7	CN128980	CN128980	RHOH1_32
45	68.8	6.6	747	7	CN124733	CN124733	RHOH1_6_C

ALIGNMENTS

RESULT 1  
AQ329770/c 888 bp DNA linear GSS 08-JAN-1999  
nbxb0045P09r CUGI Rice BAC Library Oryza sativa (japonica  
cultivar-group) genomic clone nbxb0045P09r, genomic survey  
sequence.

ACCESSION  
AQ329770 GI:4121620

VERSION  
AQ329770.1

KEYWORDS  
GSS.

SOURCE  
Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE  
1 (bases 1 to 888)  
Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)

CONTACT: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: rwing@clemson.edu  
Seq primer: GGAACAGCTATGACCATG  
Class: BAC ends  
High quality sequence start: 13  
High quality sequence stop: 225.  
location/Qualifiers

FEATURES  
source  
1..888  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbxb0045P09r"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC Library"  
/note="Vector: pBelOAC11; Site 1: HindIII; Site 2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of

carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

Query Match 14.9%; Score 155; DB 9; Length 888;  
Best Local Similarity 60.6%; Pred. No. 2.6e-25;  
Matches 251; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 150 GATTCGCGGCTATCCCTTTTGATATGCTGGCGCGCGGCTTGGGCATGCGCTTCGT 209  
DB 414 GCTAACGGGATATCCATATTAGGCATAGTAGAGATTCTAGAAATGGGGAACAGTTAGA 355  
QY 210 CCAGCGCTATTTCGAGAATTGCTGTCGGCGGACAGCAAGCAGTGGCAGCCCTGGCGGA 269  
DB 354 CGATCGACAGTCAGAAAAACAGCANGNCCGACAGATGCCCGAGTCCGCCGCGATCGCCGA 295  
QY 270 TGGCGCGCGCGCGCCACGGCATGATGCTGGCGCGCTATAGCGAGCGCGCGCGCAG 329  
DB 294 GGCGCGCGCTCGGACCGGCATCTGATCAGTCTGGCTATAGCGAGCGCGCGCGCAG 235  
QY 330 CCTCTATATGGGCGCAGCGCATCTTCGGCCCCGATGGCGATCGATCGCGCGCGCGCA 389  
DB 234 CCTCTACATCGCTCAACTACTGATCGATGATCGTGGCAGCATCGTACAGCGCGCGCA 175  
QY 390 GCTCAAGCCTACCCTATCGGAGCGCACCGTGTTCGGCGAGGAGACGGACCATCTCCG 449  
DB 174 GCTCAAGGCAACCCATGTGAGCGGACCATCTTCGGTGAGGGCGGACGCGACATCCG 115  
QY 450 GGTGCACGATACCGCCATCGGGCGGCTTCGGCGCGCTCTGTGCTGCGGAGCACATCCAGCC 509  
DB 114 CGTGTGTGAGCGGACCTTGCGCGGCTTGGATCGCTTGTGCTGCTGGGAACATATGAACCC 55  
QY 510 ATTGTGCAATATACGCCATGTACGCGCGCGGACGAAACAGGTCCACGTCGCGTGTG 563  
DB 54 GCTCACCAATATATGCCATGTACGCGCGGAGGACGACAGATCCATCTCGCGGCATG 1

RESULT 2  
LOCUS AQ329769 622 bp DNA linear GSS 08-JAN-1999  
DEFINITION nbxb0045P09f CUGI Rice BAC Library *Oryza sativa* (japonica cultivar-group) genomic clone nbxb0045P09f, genomic survey sequence.  
ACCESSION AQ329769 GI:4121619  
VERSION AQ329769.1 GI:4121619  
KEYWORDS GSS.  
SOURCE *Oryza sativa* (japonica cultivar-group)  
ORGANISM *Oryza sativa* (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; *Oryza*.  
REFERENCE 1 (bases 1 to 622)  
AUTHORS Wing, R.A. and Dean, R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence stop: 423.  
location/Qualifiers

FEATURES

source

1..622  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbxb0045P09f"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC Library"  
/note="Vector: pBelobAC11; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

Query Match 14.7%; Score 153; DB 9; Length 622;  
Best Local Similarity 60.9%; Pred. No. 7.5e-25;  
Matches 249; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

QY 577 CTCTATCGCGGCGATGGCTATGCGCTCGACCGGAGGTCAATACCGCGCAAGCCAGATC 636  
DB 5 CTCTACAACGGCAAGGCTTTCGCGCTCGGTCCGAGCTCAACAATGCCGTGACGACGTC 64  
QY 637 TACGCGGTGAGGGCGGCTGCTACGTGCTGGCGTGTGCGGACCGGTTTTCGCGGAGATG 696  
DB 65 TACGCGCGGAGGACAAATGCTTCGTATCGCGCCCTGCGGGGTGTTTCTCGGGAATG 124  
QY 697 ATCAAGTATTGTGATACGCCCGACAAAGAGATGTTCTCAAGCGCGGCGGTTT 756  
DB 125 ATCAGCTGATGTGCGATACCGAGAGCAGCGCACTTGCTGCGCGGCGCGGCTAT 184  
QY 757 GCCATGATTTTCGGCGCGGACGGCGCGCTTGCGCGGACGCGCTCCCGAGACCGAAGAG 816  
DB 185 GCGATGATCTATGCTCCGATGGCGCGCGGATGGCGGAACCAATCGCACTGAGAGGAA 244  
QY 817 GGACTGCTGTGCGCGCATATCGACCTCGGCATGATCGCGTTGGCCAAAGCGCGCGCAT 876  
DB 245 GGCTCGTATGTGCGGACATCGATCTGTGATGATTTGCCCTACGCCAAGACCGCGGACAG 304  
QY 877 CCGCGCGGCGCACTATTCACGGCGCGGACGTAACGCGGCTGCTGTGATGACGTCGCGCC 936  
DB 305 CCGACCGACATTAATGCGGACGACGATGCGCAGCGCTGCTTTCAACCGCGGACCGGACC 364  
QY 937 CAACGCGTGTGACGCTGATGCGGCATTCGAACGCCAAAACGAGACA 985  
DB 365 GCCCGGCTGAGCGGCTTATCGCGCGTCAAGCCAGCGGACCATCGCA 413

RESULT 3  
LOCUS DR631467 699 bp mRNA linear EST 11-JUL-2005  
DEFINITION EST1021595 FvI *Gibberella moniliformis* cDNA clone FVIE956, mRNA

sequence.  
ACCESSION DR631467 GI:70706307  
VERSION DR631467.1  
KEYWORDS EST.  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
REFERENCE 1 (bases 1 to 699)  
AUTHORS Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Uterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.  
TITLE Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster  
JOURNAL Unpublished (2005)  
COMMENT Contact: Brown, D.W.  
USDA/ARS/NCAUR  
USDA  
1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncaur.usda.gov  
FVIB956TH  
TIGR sequence name: FVIB956TH  
Seq primer: AAT TAA CCC TCA AAG GG.  
Location/Qualifiers  
1. 699  
/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="m3125"  
/db\_xref="taxon:117187"  
/clone="FVIB956"  
/cissue\_type="mycelia"  
/clone\_lib="FVI"  
/note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FVI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (cDNA Synthesis Kit; Stratagene)."  
ORIGIN  
Query Match 13.5%; Score 140.8; DB 8; Length 699;  
Best Local Similarity 54.2%; Pred. No. 5e-22;  
Matches 308; Conservative 0; Mismatches 257; Indels 3; Gaps 1;  
QY 9 GCCCATGACGAATATCGCGCGCGGTCAGGCGCGCGGTTCCTGATCTGA 68  
DB 53 GGCTATCACCAAGTACAGGCGCGCTGTGTCACTCCGAGCCTGATGTTGATCTTGA 112  
QY 69 CCGACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCCAGACGAGACGTGCGCT 128  
DB 113 GGGCGGTTCGCAAGACCATGACTTCATCAACGAGCGCGCCAGCTGGCTGCAAACT 172  
QY 129 GATCGCATCCAGAGACTTGATCCCGGCTATCCCTTTGATATGCTGGCGCGCC 188  
DB 173 CGTCGCCCTTCGGAAGTTTGATCCCTGGATATCCTTACTGATGTGAAGTACCTA 232  
QY 189 GGCTGGGGCATGCGCTTCGCCAGCGCTATTTCGAGAATTCCGTCGTGCGCGGACGAA 248  
DB 233 CCTCCAATCCCTCCCATGCTGAAGCGCTACCGGAGAACTCCATGCGCGCTGACTCTGA 292  
QY 249 GCACTGGCAGGCCCTGGCGGATCGCGCGCGCGCCAGCGCATGATGTCGCGGCTA 308  
DB 293 GGAATGCGCGGTATTGTCGCGCAGCGCGCGATACCAAGATCTACGTCTCCCTCGGCTT 352

QY 309 TAGCGACGCGCGCGGCGAGCTCTATATGGCCAGCGCATCTTCGCCCCGATGCGCA 368  
DB 353 CTCTGAGATTGACCAACGCAACTTTTAACTTCGCCCAAGTCTCATCGCCCCGATGGCTC 412  
QY 369 TCTGATCGCCGCGCGCGCAAGCTCAAGCTTACCATGCGAGCGCACCGTGTTCGCGCA 428  
DB 413 GGTCAACAACCAACCGTCGCAAGATCAAGCCAACCTACGTTGAGAAGCTTGATACGGCGA 472  
QY 429 GGGAGACGG---CAGCCATCTCGCGGTGCACGATACCGCCATCGCGCGCTCGCGCGCT 485  
DB 473 TGGCTCCGGGATACCTTCATGCGCGTTAGCGAGACTGACATCGCGCGTGTGGCCAGCT 532  
QY 486 CTGTTGCTGGGAGACATCCAGCCATTGTGGAATATACGCCATGACGCCCGCAGCAACA 545  
DB 533 TAACTGCTGGGAGAACATGAAACCCCTTCTCAAGTCTTCAACGTTTCTGTGTGAGCA 592  
QY 546 GGTCCAGTCGCGTCTGTGCGCGGAGCTTC 573  
DB 593 GGTTCACATCGCTGCTTGCGCCGTTCTAC 620

RESULT 4  
DR631867  
LOCUS 732 bp mRNA linear EST 11-JUL-2005  
DEFINITION EST1021995 FVI Gibberella moniliformis cDNA clone FVIB80, mRNA sequence.  
ACCESSION DR631867  
VERSION DR631867.1 GI:70706707  
KEYWORDS EST.  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
REFERENCE 1 (bases 1 to 732)  
AUTHORS Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Uterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.  
TITLE Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster  
JOURNAL Unpublished (2005)  
COMMENT Contact: Brown, D.W.  
USDA/ARS/NCAUR  
USDA  
1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncaur.usda.gov  
FVIB80TH  
TIGR sequence name: FVIB80TH  
Seq primer: AAT TAA CCC TCA AAG GG.  
Location/Qualifiers  
1. 732  
/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="m3125"  
/db\_xref="taxon:117187"  
/clone="FVIB80"  
/cissue\_type="mycelia"  
/clone\_lib="FVI"  
/note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FVI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (cDNA Synthesis Kit; Stratagene)."  
ORIGIN





LOCUS	DR632130	841 bp	mRNA	linear	EST 11-JUL-2005
DEFINITION	EST1022258 Fv1 <i>Gibberella moniliformis</i> cDNA clone FVIE179, mRNA sequence.				
ACCESSION	DR632130				
VERSION	DR632130.1				
KEYWORDS	GI:70706970				
SOURCE	EST.				
ORGANISM	<i>Gibberella moniliformis</i> <i>Gibberella moniliformis</i> Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; <i>Gibberella</i> .				
REFERENCE	1 (bases 1 to 841)				
AUTHORS	Brown,D.W., Cheung,F., Proctor,R.H., Butcho,A.E., Zheng,L, Lee,Y., Uterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.				
TITLE	Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster				
JOURNAL	Unpublished (2005)				
COMMENT	Contact: Brown, D.W. USDA/ARS/NCAUR				
	USDA				
	1815 N. University St, Peoria, IL 61604, USA				
	Tel: 309 681 6230				
	Fax: 309 681 6689				
	Email: brown@ncaur.usda.gov				
	TIGR sequence name: FVIE179TH				
	Seq primer: AAT TAA CCC TCA AAG GG.				
FEATURES	Location/Qualifiers				

```

source
1. .841
/organism="Gibberella moniliformis"
/mol_type="mRNA"
/strain="m3125"
/db_xref="taxon:117187"
/clone="FVIE179"
/tissue_type="mycelia"
/clone_lib="FV1"
/note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI;
Site 2: XhoI; anamorph: Fusarium verticillioides. Library
FVI was prepared from growth on excised maize seedling
roots and shoots. The roots and shoots were prepared as
described above and then inoculated by dipping briefly in
a suspension of 5 x 106 conidia per ml. The inoculated
roots and shoots were then incubated on moistened Whatman
#1 filter paper at room temperature under sterile
conditions for 4 days. Fungal mycelial was collected,
frozen in liquid nitrogen, ground to a powder, and then
added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
approximately 1 g mycelia per 10 ml TRIzol. The cDNA was
directionally ligated into the pBluescript II SK(+) XR
vector (cDNA Synthesis Kit; Stratagene)."

```

ORIGIN			
Query Match	13.5%;	Score 140.8;	DB 8;
Best Local Similarity	54.2%;	Pred. No. 5e-22;	length 841;
Matches 308;	Conservative	0;	Mismatches 257;
			Indels 3;
			Gaps 1;

QY 9 GCCCATGACGAAGTATCGCGCGCGGGCGGTGCAGGCCCGCCGGGTGTTCTTCGATTCGA 68  
| | | | | | | | | | | | | | | | | | | | |  
Db 32 GGCTATCACCAGAAGTACAAGGCCGCTGCTGTCACTCCGAGCTTGATGTTGATCTTGA 91  
| | | | | | | | | | | | | | | | | | | | |  
QY 69 CCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGACGTGCGCT 128  
| | | | | | | | | | | | | | | | | | | | |  
Db 92 GGGCGGTGTTGCAAGACCATCGACTTCATCAACGAGCGCGGCCAAGCTGCTGCAACT 151  
| | | | | | | | | | | | | | | | | | | | |  
QY 129 GATCGCATTTCCAGAGACTTGATTCCCGGCTATCCCCTTTTGATATGCGTGGCGGCC 188  
| | | | | | | | | | | | | | | | | | | | |  
Db 152 CGTCGCCCTTCCCGAAGTTGGATCCCTGGATATCCTTACTGATGTGGAAGGTCACCTA 211  
| | | | | | | | | | | | | | | | | | | | |  
QY 189 GGCTTTGGGGCATGCGCTTCGTCAGCGGCTATTTGAGAAATTGCTCGTGCGCGGACGAA 248  
| | | | | | | | | | | | | | | | | | | | |  
Db 212 CCTCCAATCCCTCCCCCATGCTGAAGCGCTAACCGGAGAACTCCATGCGCGTGACTTGA 271  
| | | | | | | | | | | | | | | | | | | | |  
QY 249 GCAGTGGCAGGCCCTGGCGGATGCGGCCCGCGCCACGCGCATGTGCTGCGCGGCTA 308  
| | | | | | | | | | | | | | | | | | | | |

Db 272 GGAATGCGCCGTAATTGTCGCGCAGACCCCGGATAACCAAGATCTACGTCCTCCGCTT 331

QY 309 TAGCAGCGCGCGGGCGGCAGCCTCTATATGGCCAGGCGATCTTCGCCCCGATGGCGA 368

Db 332 CTCTGAGATTGACCAAGCAACTCTTTACTTCGCCCAAGTCCTCATCGGCCCGATGGCTC 391

QY 369 TCTGATCGCGCGCGCCGCAAGCTCAAGCCTAACCCATGCGGAGCGCACCGTGTTCGCGA 428

Db 392 GGTCAATCAACCAACCGTCGCAAGATCAAGCCAACTCAAGTTGAGAAGCTGTATACGGCGA 451

QY 429 GGGAGACGG---CAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTTCGGCGCGCT 485

Db 452 TGGCTCCGGGGAATCCTTCATGCGCGTTAGCGAGACTGACATCGGCCGTGTGGCCAGCT 511

QY 486 CTGTTGCTGGAGACATCCAGCCATTGTGAAATAAGCCATGTACCGCGCGAGCAACA 545

Db 512 TAACTGCTGGAGAACATGAACCCCTTCCTCAAGTCTCTCAACGTTTCTGCTGTGAGCA 571

QY 546 GGTCCACGTCGCGTCGTGGCCGAGCTTC 573

Db 572 GGTTCACATCGCTGCTTGGCCCCGTTAC 599

RESULT 7	
DR622046	
LOCUS	DR622046
DEFINITION	EST1012174 Fv1 Gibberella moniliformis cDNA clone FVIA704, mRNA
	856 bp
	mRNA
	linear
	EST 11-JUL-2005

ACCESSION	sequence.
VERSION	DR622046
KEYWORDS	DR622046.1
SOURCE	EST.
ORGANISM	Gibberella moniliformis
	Gibberella moniliformis

REFERENCE  
1 (bases 1 to 856)  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

**AUTHORS** Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L, Lee, Y.,

TITLE	Author
Utterback, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D., Kendra, D.F., Town, C.D. and Whitelaw, C.A.	Analysis of 87,000 expressed sequence tags reveals alternatively

JOURNAL  
Unpublished (2005)

**COMMENT**  
**Contact: BROWN, D.W.**  
**USDA/ARS/NCAUR**

USDA  
1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncaur.usda.gov  
TIGR sequence name: FVIA704TH  
Seq primer: AAT TAA CCC TCA AAG GG.

```

FEATURES
    source
        1. 856
            location/Qualifiers
            /organism="Gibberella moniliformis"
            /mol_type="mRNA"
            /strain="m3125"
            /db_xref="taxon:117187"
            /clone="FVIA704"
            /tissue_type="mycelia"
            /clone_lib="FVI"
            /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI;
            Site_2: XhoI; anamorph: Fusarium verticillioides. Library
            FVI was prepared from growth on excised maize seedling
            roots and shoots. The roots and shoots were prepared as
            described above and then inoculated by dipping briefly in
            a suspension of 5 x 106 conidia per ml. The inoculated
            roots and shoots were then incubated on moistened Whatman
            #1 filter paper at room temperature under sterile
            conditions for 4 days. Fungal mycelial was collected,
            frozen in liquid nitrogen, ground to a powder, and then
            added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
            approximately 1 g mycelia per 10 ml TRIzol. The CDNA was
            directionally ligated into the pBluescript II SK(+) XR

```





```
RESULT 9
CF453420          655 bp      mRNA      linear      EST 01-JAN-2004
LOCUS             ft536_1124 Fusarium verticillioides fec1 mutant subtraction
DEFINITION       library Gibberella moniliformis cDNA, mRNA sequence.
ACCESSION        CF453420
VERSION          CF453420.1  GI:40546865
KEYWORDS
SOURCE           EST.
ORGANISM         Gibberella moniliformis
                  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
                  Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE        1 (bases 1 to 655)
AUTHORS          Pirttila,A.M., McIntyre,L.M., Payne,G.A. and Woloshuk,C.P.
TITLE            Comparison of gene expression in the wild type and the fec1 mutant
                  of Fusarium verticillioides reveals expressed sequence tags
                  associated with fumonisin biosynthesis
JOURNAL          Unpublished (2003)
COMMENT          Contact: Woloshuk CP
                  Department of Botany and Plant Pathology
                  Purdue University
                  915W. State Street, West Lafayette IN 47907-2054, USA
                  Tel: 765 494 3450
                  Fax: 765 494 0363
                  Email: woloshuk@purdue.edu
                  Insert Length: 655 Std Error: 0.00
                  Plate: ft row: 0 column: 11
                  Seq primer: T7.
FEATURES
  source          location/Qualifiers
                  1..655
                  /organism="Gibberella moniliformis"
                  /mol_type="mRNA"
                  /strain="FT536"
                  /db_xref="taxon:117187"
                  /clone_lib="Fusarium verticillioides fec1 mutant
                  subtraction library"
                  /note="Vector: Pgem-T-Easy; Site_1: EcoRI; Site_2: EcoRI;
                  Fungus was grown on cracked corn. RNA was isolated using
                  phenol LiCl method. Polya RNA was obtained with Oligotex
                  mRNA spin columns (Qiagen). Subtracted from wild type RNA
                  with PCR select cDNA subtraction kit (Clontech),
                  amplified, cloned into pGEM-TEZ and transformed to E. coli
                  DH5 alpha cells."
ORIGIN
Query Match      12.5%; Score 129.8; DB 6; Length 655;
Best Local Similarity 53.6%; Pred. No. 1.8e-19;
Matches 294; Conservative 0; Mismatches 252; Indels 3; Gaps 1;

QY      28 GGGCGGGCGGTGACAGCGCGCGCGGGTGTCTCTGATCTGACCGCACAGTGGAGAAAGCG 87
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7 GCGCGTGTGTACCTCGAGCCTGAGTGTTCGATCTTGAGGGCGGTGTTGCAAGACC 66

QY      88 ATCGGCCTGATCGAGCAGCGCGCCCAAGCAGACGTGCGCTGATCGCATTCCAGAGACT 147
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      67 ATCGACTTCATCAACGAGCGCGGCCAAGCTGCTGCAAACTCGTCGCTTCCCGAAGTT 126

QY     148 TTGATTCGCCGCTATCCCTTTTGATATAGCTGGCGCGCGCGCTTGGGCATGCGCTTC 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     127 TTGATCCCTGATATCTTA CTGATGTGGAAGTCACTCACTCCAATCCCTCCCATG 186

QY     208 GTCCAGCGCTATTTCGAGAAATTCGCTCGTGGCGGCGAGCAGCAGTGGCAGGCGCTGGCG 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     187 CTGAGGCGCTACCGCGAGAACTCCATGCGCGTCACTCTGAGGAATGCGCGGTTCGT 246

QY     268 GATCGGGCGCGCGCCACGGCATGCATGCTGTGCGCGGCTATAGCGAGCGCGGGGCGGC 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     247 CGCGCAGCGCGCGATTAACAGATCTACGTCTCCCTCGGCTTCTGTAGATTGACCAAGCA 306

QY     328 AGCCTTATATGGGCGCAGCGCATCTTGGCGCGCGGATGCGATCGCGCGCGCGCGC 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     307 ACTCTTACCTGCGCCAAAGTCTCATCAGCCCCGATGGCTCGGTCAACCAACCGTTCG 366
```

```
QY      388 AAGCTCAAGCCTTACCCTATCGGAGCGCACCGTGTTCGGCGAGGAGACGG---CAGCCAT 444
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      367 AAGATCAAGCCAACTACGTTGAGAAGCTTGATATACGGCGATGCTCCGGGATACCTTC 426

QY      445 CTCGGGTGACAGTACCCTATCGGCGCGCGCTGTGTGCTGGGAGCACATC 504
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      427 ATGCCGTTAGCGAGACTGACATCGCGCGTGTGGCCAGCTTAAGTCTGGGAGAACATG 486

QY      505 CAGCCATTGTGGAATAAGCCATGTACGCCCGCGCAGACAGAGTCCACGTGCGTCTGG 564
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      487 AACCCTTCTCAAGTCTCTCAACGTTTCTGTGATGAGCAGGTTCAATCGCTGCTTGG 546

QY      565 CCGAGCTTC 573
      ||| ||| |||
Db      547 CCGGTCTAC 555

RESULT 10
CN808140          840 bp      mRNA      linear      EST 27-MAY-2004
LOCUS             Blood EST0223 Metarhizium anisopliae ARSEF 2575 from insect blood
DEFINITION       Metarhizium anisopliae cDNA clone B602 5', mRNA sequence.
ACCESSION        CN808140
VERSION          CN808140.1  GI:47729613
KEYWORDS
SOURCE           EST.
ORGANISM         Metarhizium anisopliae
                  Metarhizium anisopliae
                  Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
                  Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
                  Clavicipitaceae; Metarhizium.
REFERENCE        1 (bases 1 to 840)
AUTHORS          Wang,C.S., Hu,G. and St. Leger,R.J.
TITLE            Gene expression profiling of Metarhizium anisopliae grown under
                  different conditions: mechanisms of fungal opportunism
JOURNAL          Unpublished (2004)
COMMENT          Contact: Wang CS
                  Department of Entomology
                  University of Maryland
                  4112 Plant Sciences Building, College Park, MD 20742, USA
                  Email: cwang4@umd.edu
                  Seq primer: M13 Reverse.
FEATURES
  source          location/Qualifiers
                  1..840
                  /organism="Metarhizium anisopliae"
                  /mol_type="mRNA"
                  /db_xref="taxon:5530"
                  /clone_lib="B602"
                  /clone_lib="Metarhizium anisopliae ARSEF 2575 from insect
                  blood"
                  /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was
                  grown in insect haemolymph for 24 hours. A cDNA library
                  was constructed in the vector pCMV.SPORT6.1"
ORIGIN
Query Match      12.4%; Score 129.4; DB 7; Length 840;
Best Local Similarity 53.1%; Pred. No. 2.2e-19;
Matches 299; Conservative 0; Mismatches 261; Indels 3; Gaps 1;

QY      20 AGTATCGGGCGGGGCTGACAGCGCGCGGGTTCCTCGATCTGACCGCAGACTCG 79
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      80 AGTACAAGGCTGTGTGTGCAAGCTGAGCTTGATGGCTCAACCTTGAGAAAGCAGTCA 139

QY      80 AGAAAGCGATCGGCTGATCGAGCAGCGCGCCAAAGCAGACGTCGCGCTGATCGATTCC 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     140 AGAAGACTGTGACTTGATCTCTCGAAGCTGGCGAGAAAGATTGCAAGCTTATTCCTTTC 199

QY     140 CAGAGACTTGATTCGCGGCTATCCCTTTTGAATATGAGCTGGCGCGCGCTTGGGGCA 199
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     200 CTGAAGTTTGATTCCTGGAATTCTTCCATTGGCTGTGGCGTGTCAACTACCAAGATTCCG 259

QY     200 TGGCTTCGTCACGCGTATTTGAGAATTGCTCGTGGCGGCGCAGCAAGCAGTGGCAGG 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     260 TGCCCTCTCAAGAGTTTCAACGAAACAGACGATGCGGCCGACTCGAGCAATGCGAC 319
```



[illegible]

RESULT 11  
LOCUS CN808249  
DEFINITION Blood EST0332 Metarhizium anisopliae ARSEF 2575 from insect blood  
Metarhizium anisopliae CDNA clone B711 5', mRNA sequence.  
ACCESSION CN808249  
VERSION CN808249.1 GI:47729722  
KEYWORDS EST.  
SOURCE Metarhizium anisopliae  
ORGANISM Metarhizium anisopliae  
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic  
Clavicipitaceae; Metarhizium.  
1 (bases 1 to 998)  
AUTHORS Wang, C.S., Hu, G. and St. Leger, R.J.  
TITLE Gene expression profiling of Metarhizium anisopliae grown under  
different conditions: mechanisms of fungal opportunism  
JOURNAL Unpublished (2004)  
COMMENT Contact: Wang CS

```

FEATURES
    source
        Location/Qualifiers
            1..998
                /organism="Metarhizium anisopliae"
                /mol_type="mRNA"
                /db_xref="Caxon:5530"
                /clone="B711"
                /clone_1lb="Metarhizium anisopliae ARSEF 2575 from insect
                blood"
                /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was
                grown in insect haemolymph for 24 hours. A cDNA library
                was constructed in the vector pCMV.SPORT6.1"

```

Query Match	12.4%;	Score 129.4;	DB 7;	Length 998;
Best Local Similarity	53.1%;	Pred. No. 2.2e-19;		
Matches 299; Conservative	0;	Mismatches 261;	Indels 3;	Gaps 1;

QY 20 AGTATCGCGCGCGGTGCAGGCCGCCGGTGTTCTTCGATCTCGACCGCACAGTGC 79  
 |||| | | | | | | | | | | | | | | | | | |  
 Db 104 AGTACAAGGCTGCTTGTTGTGTGCAGCTGAGCCTGATGCTTCACCTTGAGAAAGCAGTCA 163  
 | | | | | | | | | | | | | | | | | | | | | |  
 QY 80 AGAAAGCGATCGGCTGATCGAGCAGGCGGCCAAGCAGGACGTGCGCCTGATCGCATTC 139  
 |||| | | | | | | | | | | | | | | | | | |

Db	164	AGAAGACTGTTGACTTGATTCCTCGAAGCTGGCGGAGAAAGATTGCAAGCTTATTGCTTTC	223
QY	140	CAGAGACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGCGGCTTGGGCA	199
Db	224	CTGAAGTTGGATTCTTGGAFTTCCCTATTGGCTGTGGCGTGTCACTACCAAGATTGCG	283
QY	200	TGCGCTTCGTCCAGCGCTAFTTCGAAATTGCTCGTGGCGGGCAGCAAGCAGTGGCAG	259
Db	284	TGCCCCCTCTCAAGAAATTTCACCAAGAACAGCATGGGCGCGCACTCGGACGAAATGCGAC	343
QY	260	CCCTGGCGGATGCGGCGCGCGCCACAGGCATGATGTCTGTGGCCGGCTATAGCAGCGG	319
Db	344	GAATTGCGGAGGCTGCCAAAGCCGCGTAAGATCTACGTCGTGGGATATTCCGAGCTCG	403
QY	320	CGGGCGGACGCTCTAATATGGGCCAGGCGGATCTTCGGCGCCGATGGCGATTGATCGCG	379
Db	404	ACGGCCATACCATGTACATGGCGCAGATCATCATTTGATTCACCGGAACCGTCATCAACC	463
QY	380	CGCGCCGCAAGCTCAAAGCTACCCCATGCGGAGCGCACCGTGTTCGGCGAGGGAGACGG--	437
Db	464	ACCGTCGCAAGATCAAGCCCAACCCATGTGGAAGCTTGTATTGGCGAGGGAGCGGTG	523
QY	438	-CAGCCATCTCGCGGTGCAAGATACCGCCATCGGCGGCTCGGCGCGCTGTGTGCTGG	496
Db	524	ACTCTCTTACAGCCGCTAGTGAAGACTGAATTTGGCAATCTCGGCCACCTGAACCTGCTGG	583
QY	497	AGCACATCCAGCCATTTGTGAATAACGCCATGTACGCGCGCGCAGCAACAGTCCACGTCG	556
Db	584	AAAATATGAACCCCTTCTCTCAAGGCGCTCAACGCCAGCAAAACAGAAATAATCCACGTAG	643
QY	557	CGTCGTGGCCGAGCTTCAGCCTC	579
Db	644	CCGCATGGCCCGTGTACGGCCCC	666

RESULT	12
DR609823	
LOCUS	
DEFINITION	769 bp mRNA linear EST 11-JUL-2005 EST999951 FVG Gibberella moniliformis cDNA clone FVGBB83, mRNA sequence.
ACCESSION	DR609823
VERSION	DR609823
KEYWORDS	DR609823.1 GI:70684471
SOURCE	EST.
ORGANISM	Gibberella moniliformis Gibberella moniliformis Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE	1 (bases 1 to 769)
AUTHORS	Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Uterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.
TITLE	Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster unpublished (2005)
JOURNAL	Contact: Brown, D.W.
COMMENT	USDA/ARS/NCAUR

```

FEATURES
    source
        Location/Qualifiers
            1. 769
                /organism="Gibberella moniliformis"
                /mol_type="mRNA"
                /strain="m3125"
                /db_xref="taxon:117187"
                /clone="FVGDB83"
                /tissue_type="mycelia"
                /clone_lib="FVG"
                /note="Vector: pBluescript II SK(+) XR; Site_1: EcoRI;

```

Site 2: XhoI; anamorph: Fusarium verticillioides. Mycelia was collected after growth in liquid GYM medium for 96 hours. Cultures were vacuum filtered and the mycelial mats were frozen in liquid nitrogen, ground to a powder, and then added to TRIzol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml TRIzol. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (cDNA synthesis kit; Stratagene)."

ORIGIN

Query Match 12.3%; Score 127.6; DB 8; Length 769;  
Best Local Similarity 51.5%; Pred. No. 5.7e-19;  
Matches 348; Conservative 0; Mismatches 319; Indels 9; Gaps 2;

OY	28	GGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTGACCGCACAGTCGAGAAAGCG	87
DB	61	GTCGCCGCATCCAAAGCCGAGCCCGTGGAAAGATCTCCAGGGTGGCGTCAACAAGTCC	120
OY	88	ATCGGCTGATCGAGCAGCGCGCAAGCAGACGTGCGCTGATCGATTCCAGAGACT	147
DB	121	ATCGGTCTCATCCAAGAGCGACAAAGAGGGTGCCAAAGTCATGGCTTCCCTGAAGTC	180
OY	148	TGGATTCGCGGCTATCCCTTTTGATATGGCTGGCGCGCGCTGGGCATGCGCTTC	207
DB	181	TTCATCCCTGGATATCCATGAGCATCTGGGCCAACTCGCCTACCGAAGACGACCATGG	240
OY	208	GTCCAGCGCTATTTCGAGAATTGCTCGTGGCGCGGACGAAGCAGTGGCAGGCCCTGGCG	267
DB	241	ATCAATGAGTACTTCAAGAACTCAATGAGAAAGATCACCCTGATGAGCCAGATCCGA	300
OY	268	GATGCGCGCGCGCGCCACGCGCATGTCGTGGCGCGCTATAGCGAGCGCGCGCGGC	327
DB	301	GCTGCTGTTGAGAGGAGGAGGTGTTTGTAGTCTTGGATACAGCGAGATACAGGGGA	360
OY	328	AGCCTTAATGAGGCCAGGCGATCTTGGCGCGCGATGGCGATCGCGCGCGCGGC	387
DB	361	ACCCTTACATCGCACAGTCTTTCATCGAGAGACCGGCATATTGTTCCACCGTCGC	420
OY	388	AAGCTCAAGCCTAACCCATGCGGAGCGACCGTGTTCGGGAGGAGACGGCAGCATCT-	446
DB	421	AAGATCAAGCCTAACCCAGTGAAGCGTGTATCTACGGTGACGGAGCGGCGAGTCTTG	480
OY	447	--CGGGTGACGATACCGGCATCGGGCGCTCGGGCGCTGTGTTGCTGGAGCACATC	504
DB	481	ACCAACGTGCGCGACACAAGTTCGGAAGGTTGCTGTTAACTGTGGGAGCACACC	540
OY	505	CAGCCATTGTGAATAAGCCCATGTACCGCGCGGACGAACAGGTCCAGTCCGCTGGG	564
DB	541	CAGACACTTCTCCGCTACTACGAATACCTCCAGACGTGATATCCAGTCTCCAGCTGG	600
OY	565	CCGAGCTTACGCTCTA-----TCGGCGCATGGCCTATGCGTCCGACCGGAGTCAAT	618
DB	601	CCTTCTATCTTCCCCAGAACGTCCCGAGTGGCCATACCAACATCACTCCGAATGCTGC	660
OY	619	ACCGCGCAAGCAGATCTACGGGTGAGGGCGGCTGCTACGTGTCGCGCTGTCGCGG	678
DB	661	AAGGCTTCTCTACGTCGTCTCCATGGAAGAGCCTGCTTCGTTCTCTGGAAGTCAAG	720
OY	679	ACCGTTTCGCGGAGGA 694	
DB	721	ATCATGACTGAGGAGA 736	

RESULT 13

LOCUS	DR709497	840 bp	mRNA	linear	EST 14-JUL-2005
DEFINITION	Asn_11063 Aspergillus niger pBluescript (EcoRI-XhoI) Aspergillus niger cDNA clone Asn_11063, mRNA sequence.				
ACCESSION	DR709497				
VERSION	DR709497.1	GI: 70825788			
KEYWORDS	EST.				
SOURCE	Aspergillus niger				
ORGANISM	Aspergillus niger				

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus. 1 (bases 1 to 840)  
AUTHORS Tsang,A., Storms,R. and Bulter,G.  
TITLE Expressed sequence tags from Aspergillus niger cDNA library  
JOURNAL Unpublished (2005)  
COMMENT Contact: Tsang A  
Centre for Structural and Functional Genomics  
Concordia University  
7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
Tel: 514 848 2424 3405  
Fax: 514 848 4504  
Email: tsang@vax2.concordia.ca  
POLYA=No.

FEATURES

location/Qualifiers  
1..840  
/organism="Aspergillus niger"  
/mol\_type="mRNA"  
/strain="N402"  
/db\_xref="taxon:5061"  
/clone="Asn\_11063"  
/dev\_stage="mycelial growth"  
/lab\_host="E. coli"  
/clone\_11b="Aspergillus niger pBluescript (EcoRI-XhoI)"  
/note="Vector: pBluescript KS+; Site\_1: XhoI; Site\_2: EcoRI; Complementary DNA was synthesized with ZAP kit (Stratagene) using poly(A)+RNA isolated from Aspergillus niger cultured under different carbon sources (glucose, maltose, xylose, lactose, sorbitol, xylan, and bran). Synthesis was primed with oligo(dT)/XhoI primer. EcoRI adaptors were ligated to the blunt-ended, double-stranded cDNA. The EcoRI-XhoI-digested cDNA was ligated with EcoRI-XhoI-digested pBluescript KS+ (Invitrogen Corp)."

ORIGIN

Query Match 11.8%; Score 122.6; DB 8; Length 840;  
Best Local Similarity 50.3%; Pred. No. 8.2e-18;  
Matches 417; Conservative 0; Mismatches 394; Indels 18; Gaps 4;

OY	25	CGCGCGCGCGGTGCAGGCGCGCGGTGTTCTCGATCTGACCGCACAGTCGAGAAA	84
DB	19	CGCGTCGAGTACCCAGGACAGCTGTGCTGACTTGGAAGCAACCGTGAAGGAAAG	78
OY	85	GCGATCGGCTGATCGAGCAGCGCGCAAGCAGACGTGCGCTGATCGATTCCAGAG	144
DB	79	ACATGTAACCTTATTGCTGAAGCAGCTCGAAGCGGCTCAGTTGGTGAATTTCCGAG	138
OY	145	ACTTGAATCCCGGCTATCCCTTTGATATGGCTGGGCGCGCGCTTGGGGCATGCGC	204
DB	139	TGTTGATCCCGGATATCCTGCTGATTTGG-----GCACGGCTGTTGACATGCGC	192
OY	205	TTGCTCAGCGCTATTTTCGAAATTCGCTCGCGCGGACGAAGTGGAGAGCCCTG	264
DB	193	CTATCATCTATTTCATACAAACTCCCTGAAGATTGATTCGCCGAATGGCCAGCATC	252
OY	265	GCGATGCGCGCGCGCGCACCGCATGTCATGTCGTCGCGCGGCTATAGCGAGCGCGGC	324
DB	253	CAGCAATGCGCAGCGGAGATTAATCGTGTGTTGGCTTTTCCGAGAACCTGCAT	312
OY	325	GGCAGCTTATATGGGCCAGCGCATCTTGGCGCGCGGATGGCAATGATCGCGCGCGC	384
DB	313	AACCTCCGTATATCTCGCAGCGCTATTATGCAAGTGAAGGAAAGATCCTCACACCCGC	372
OY	385	CGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTGTTCGCGA---GGAGAGCGCAGC	441
DB	373	AAGAAATCAAGCCCACTCATATGAGCGGACCATTTTCGGTGACTCGTTTGAGACTGT	432
OY	442	CATCTCGGGTGACAGATACCGCATCGGGCGCTCGGGCGCTGTGCTGGAGCAC	501
DB	433	CTGCAAGTGTAGTGAATACATCAGCCGGTGGCTGTGCTCTTCTGCTGGAGCAT	492
OY	502	ATCCAGCATTTGTGAATACGCGATGACGCGCGGACGAACAGGTCCACGTCGCGTCG	561

Db 493 ATCCAGCTCTGCTCAAGTATACACTTATGCGCAGCGGAACAATTATGTTGCCGA 552  
Qy 562 TGGCCGAGCTTCAGCCTCTATCGCGCATGGCCTA-----TGGCTCGGACCGGAGGTC 615  
Db 553 TGGCCGCTCTGTTCCTCTATAGCGAGCGGTTCTTGTTCGATGTCCACCGAGGCG 612  
Qy 616 AATACCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGTACGTGCGCTGCGTCGC 675  
Db 613 ACCAGTTCATGTCGAGGACCTATGCAAGAGTCAAGTCTTTGTCTCCATAC---C 669  
Qy 676 GCGACCGTTTCGCCGAGATGATCAAGTATGTTGTGATACGCCCGACAAGAGATGTT 735  
Db 670 ACAACAGTCATAGGCCAGTCTGTCATTCGACCGATGCGCACATCGACTGGCCCTGATG 729  
Qy 736 CTCAAGCGCGCGCGGCTTTTGCCATGATTTTTCGGCGCCGACGCGCGCCCTGGCCGAG 795  
Db 730 AGCACCCCTGGCGGAGGTTGCTCCGCCATCTTCGGCGCCGACGGTCCCAATTATCGCAA 789  
Qy 796 CCGCTCCGCGAGACCGAAGAGGAGTGTGTGCGCCGATATCGACTCG 844  
Db 790 CCATTCCTAATGACAGAGAGGGGATTAATCTACCCCGACCTCGACTTTG 838

RESULT 14  
AQ160684 580 bp DNA linear GSS 09-SEP-1998  
LOCUS mgxb0006L04r CUGI Rice Blast BAC Library Magnaporthe grisea genomic  
DEFINITION clone mgxb0006L04r, genomic survey sequence.  
ACCESSION AQ160684  
VERSION AQ160684.1 GI:3557673  
KEYWORDS GSS.  
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
ORGANISM Magnaporthe grisea  
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
AUTHORS Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.  
1 (bases 1 to 580)  
Yu.Y., Zhu,H., Boyd,C.A., Gaudette,B., Gayle,A., Kingsbury,R.,  
Phillips,K., Sasinowski,M., Wing,R.A. and Dean,R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Magnaporthe grisea  
Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Dean RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson University, Clemson, SC 29634  
Tel: 864 656 5737  
Fax: 864 656 4293  
Email: rdean@clemson.edu  
Seq primer: GGAACACGCTATGACCATG  
Class: BAC ends  
High quality sequence stop: 357.

FEATURES  
source  
1..580  
/organism="Magnaporthe grisea"  
/mol\_type="genomic DNA"  
/strain="70-15"  
/db\_xref="taxon:148305"  
/clone="mgxb0006L04r"  
/tissue\_type="Protoplasts"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice Blast BAC Library"  
/note="Vector: PBACWICH; Site 1: HindIII; Site 2: HindIII;  
Rice blast is one of the most devastating fungal diseases  
of rice world wide. It is a filamentous ascomycete with  
a haploid genome (n=7) of approximately 40 Mbp. Rice  
blast is an important model fungal pathogen for studying  
numerous aspects of the fungal-host interaction. In  
order to facilitate genome wide analysis, a BAC library  
containing 9216 clones with an average insert size of 130  
kbp was constructed. This library represents greater  
than 25X genome coverage. High density colony filters  
are available upon request."

ORIGIN

Query Match 11.2%; Score 116.6; DB 9; Length 580;  
Best Local Similarity 56.2%; Pred No. 2e-16;  
Matches 240; Conservative 0; Mismatches 184; Indels 3; Gaps 1;  
Qy 143 AGACTTGATTCCTCCGCTATCCCTTTTGATATGCTGGCGCGCGCTTGCGGCATGC 202  
Db 82 AAACAGGATTCCTGGCTACCCCTATTGATGTGAAAGTCAACTACTTGACAGACCTCC 141  
Qy 203 GCTTCGTCAGCGCTATTTCGAGAAATTCGCTCGTCGCGCGGCAAGACAGTGGCAGGCC 262  
Db 142 CGATGCTCAAGGCTTACCGCGAGAACTCTCTCAAGGTTGACTCGGACGAGATGCGCGCA 201  
Qy 263 TGGCGATGCGGCGCGCGCGCGCGCATGATGTCGTGCGCGGCTATAGCGAGCGCGG 322  
Db 202 TCCGGCGCGGCTCGCAACACATCTACGTCTCACTGGGCTTCTCCGAGATTGACC 261  
Qy 323 GCGGAGCCTCTATATGGGCGAGCGATCTTCGCGCCCGATGCGCATCTGATCGCGCGC 382  
Db 262 ACGCAACCTGTATCTGGCGAGGTGCTGATCGGCGCGGACGCGGAGTCTCAATCAC 321  
Qy 383 GCCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGTGTTCGCGGAGGAGACGG--CA 439  
Db 322 GCGCGAAGATCAAGCCCAACCATGTGAGAAAGCTTGCTATGCGCATGGCTCCGGTGATA 381  
Qy 440 GCCATCTCGCGGTGACAGTACCGCCATCGGCGCGCTCGCGCTGTGCTGGAGC 499  
Db 382 GCTTCATACCCGTGACGCAACAACACTCGCGCGCTCGGTCAACTCACTGCTGGAGA 441  
Qy 500 ACATCCAGCCATTGTGAAATAGCCCATGTACGCGCGCGCAAGACAGTCCACGTCCGCT 559  
Db 442 ACATGAATCCGCTCCTCAAGTCTGAGCGTGCCTGCGCGTGGGCGGTGCAGGTCATGTCGCG 501  
Qy 560 CGTGCC 566  
Db 502 CGTGCC 508

RESULT 15  
DR709485 804 bp mRNA linear EST 14-JUL-2005  
LOCUS Asn\_11049 Aspergillus niger pBluescript (EcoRI-XhoI) Aspergillus  
DEFINITION niger cDNA clone Asn\_11049, mRNA sequence.  
ACCESSION DR709485  
VERSION DR709485.1 GI:70825776  
KEYWORDS EST.  
SOURCE Aspergillus niger  
ORGANISM Aspergillus niger  
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
AUTHORS Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 804)  
Tsang,A., Storms,R. and Bulter,G.  
TITLE Expressed sequence tags from Aspergillus niger cDNA library  
JOURNAL Unpublished (2005)  
COMMENT Contact: Tsang A  
Centre for Structural and Functional Genomics  
Concordia University  
7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
Tel: 514 848 2424 3405  
Fax: 514 848 4504  
Email: tsang@vax2.concordia.ca  
POLYA=NO.

FEATURES  
source  
1..804  
/organism="Aspergillus niger"  
/mol\_type="mRNA"  
/strain="N402"  
/db\_xref="taxon:5061"  
/clone="Asn\_11049"  
/dev\_stage="mycelial growth"  
/lab\_host="E. coli"  
/clone\_lib="Aspergillus niger pBluescript (EcoRI-XhoI)"  
/note="Vector: pBluescript KS+; Site\_1: XhoI; Site\_2:

ECORI; Complementary DNA was synthesized with ZAP kit (Stratagene) using poly(A)+RNA isolated from Aspergillus niger cultured under different carbon sources (glucose, maltose, xylose, lactose, sorbitol, xylan, and bran). Synthesis was primed with oligo(dT)/XhoI primer. EcorI adaptors were ligated to the blunt-ended, double-stranded cDNA. The EcorI-XhoI-digested cDNA was ligated with EcorI-XhoI-digested Bluescript KS+ (Invitrogen Corp)."

## ORIGIN

```
Query Match      10.5%; Score 109.4; DB 8; Length 804;
Best Local Similarity 52.2%; Pred. No. 9.3e-15;
Matches 295; Conservative 0; Mismatches 261; Indels 9; Gaps 2;

QY      25 CCGCGCGCGCGGTGAGAGCGCGCGCGGTGTTCTCGATCTGCACCGCACAGTCGAGAAA 84
      |||||
Db      20 CGCGTCGAGTCAACCAAGCAGAGCAGCTGTGGCTGGAAGCAACCGTGAAGAAAAG 79

QY      85 GCGATCGCCTGATCGAGCAGCGCGCAAGCAGAGCTGCGCTGATCGCATTCACAGAG 144
      |||||
Db      80 ACATGTGACCTTATGTGGAAGCAGCTGCAACGCGCTCAGTTGTGACTTTCCCGAG 139

QY      145 ACTTGATTCGCGCTATCCCTTTTGATATGGCTGGCGCGCGCTTGGGCAATGCCG 204
      |||||
Db      140 TGTGGATCCCGGATATCCTGCTGATTTGG-----GCACGCGCTGTGACATGCCG 193

QY      205 TTCGTCAGCGCTATTTGAGAATTGCTCGTCCGCGCAGCAAGCAGTGGCAGGCCCTG 264
      |||||
Db      194 CTATCATCTATTATTAACAATACTCCCTGAAGATTGATTCGCCCAATGGCCAGCATC 253

QY      265 GCGATGCGGCGCGCGCGCAAGCATGTCGCGCGCTATAGCGAGCGCGCGGC 324
      |||||
Db      254 CAGCAATGCGCAGCGAGAAATAAATCGTGTGTGTGGCTTTCCGAACTTGCAT 313

QY      325 GGCAGCCTCTATATGGGCCAGCGATCTTCGCGCGCGATGGCATCTGATCGCGCGCGC 384
      |||||
Db      314 AACTCCCTGATATCTCGCAGGCTATATTGCAAGTTATGGGAAGATCCTCACACCGCG 373

QY      385 CGCAGCTCAAGCCTACCATGCGAGCGCACCGTTCGGCGAGGAGAGCGCACCAT 444
      |||||
Db      374 AAGAAATCAAGCCCACTCATATGAGCGGACCATTTCCGCTGACTCGTTGAGACTGT 433

QY      445 CTCGCG--GTGCACGATACCGCCATCGGCGCGCTCGCGCGCTGTGCTGGAGACAC 501
      |||||
Db      434 CTGCAGAGTGTGATCATATCAGCCGCTCGCTGTGCTCTTCTGCTGGAGACAT 493

QY      502 ATCCAGCCATTGTGAATAAGCCATGTACGCGCGCAGAACAGGTCCACGTGCGCTCG 561
      |||||
Db      494 ATCCAGCCTCTGTCAAGTATCACACTTATGCGCAGCGGAAACAATTCATGTTGCCGCA 553

QY      562 TGGCCGAGCTTCAGCCTCTATCGCG 586
      |||||
Db      554 TGGCCGCGCTCTGTCCCTCATAGCG 578
```

Search completed: April 27, 2006, 01:35:56  
Job time : 4905.58 secs



**THIS PAGE BLANK (USPTO)**



/protein\_id="AAR97444.1"  
/db\_xref="GI:40890199"  
/translation="MSEPMTRYKGAAYVQAPVFLDLDRIVEKAIGLIEQAAKQDVRLL  
AFPEWTIPGYFWIWLGAAPWGMRFVORYFENSLVRGSKQWQALADARRHGMHVAG  
YSERAGSLYMGQALFGPDGLIARRLKPTHAERTVFEGDGSILAVIDTAIGRLG  
ALCCWEHIQPLSKYAMYADEQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGG  
CYVLASCATVSPSEMIKVLVDTDPDKEMFLKAGGFAMI FPGDGRALAEPLPETEGLLV  
ADIDLGMIALAKAAADPAGHYSRPDVTIRLLDRRAQRVVTLDAFEPQNEKGDAPA  
LRVAESAAAAQ"

ORIGIN

Query Match	100.0%;	Score 1041;	DB 3;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.1e-124;		
Matches 1041;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGCGGTGCAGCGCGCGGTGTTCTC	60	
DB	1	ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGCGGTGCAGCGCGCGGTGTTCTC	60	
QY	61	GATCTGACCGGACAGTGCAGAAAGCGATCGGCTGATCGACGAGCGGCCAAGCAGAG	120	
DB	61	GATCTGACCGGACAGTGCAGAAAGCGATCGGCTGATCGACGAGCGGCCAAGCAGAG	120	
QY	121	GTGCGCTGATCGCATTCACAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG	180	
DB	121	GTGCGCTGATCGCATTCACAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG	180	
QY	181	GGCGCGCGGCTTGGGGCATGCGCTTCGCCAGCGCTATTTCCAGAAATTCGCTCGT	240	
DB	181	GGCGCGCGGCTTGGGGCATGCGCTTCGCCAGCGCTATTTCCAGAAATTCGCTCGT	240	
QY	241	GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGGCGCGCGCGCGCATGATGTCGTG	300	
DB	241	GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGGCGCGCGCGCGCATGATGTCGTG	300	
QY	301	GCCGCGCTATAGCAGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCG	360	
DB	301	GCCGCGCTATAGCAGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCG	360	
QY	361	GATGGCATCTGATCGCGCGCGCGCGCGCTCAAGCCTACCATGCGGAGCGCAC	420	
DB	361	GATGGCATCTGATCGCGCGCGCGCGCGCTCAAGCCTACCATGCGGAGCGCAC	420	
QY	421	TTCCGGCAGGAGAGCGGACGCCATCTCGCGGTGCAAGATACCGCCATCGGGCG	480	
DB	421	TTCCGGCAGGAGAGCGGACGCCATCTCGCGGTGCAAGATACCGCCATCGGGCG	480	
QY	481	GCGCTCTGTTGCTGGGAGCAGATCCAGCCATTGTCGAATACGCCATGTACGCC	540	
DB	481	GCGCTCTGTTGCTGGGAGCAGATCCAGCCATTGTCGAATACGCCATGTACGCC	540	
QY	541	GAAAGGTCACGTCGCGCTGCGCGGAGCTTCAGCCTCTATCGGGCATGGCTATG	600	
DB	541	GAAAGGTCACGTCGCGCTGCGCGGAGCTTCAGCCTCTATCGGGCATGGCTATG	600	
QY	601	CTCGGACCGGAGGTCATATACCGCGCAAGCCAGATCTACGCGGTGAGGGCGCT	660	
DB	601	CTCGGACCGGAGGTCATATACCGCGCAAGCCAGATCTACGCGGTGAGGGCGCT	660	
QY	661	GTGCTGACGTCGTGCGGACCGTTTCGCGGAGATGATCAAGGTAATTGATACG	720	
DB	661	GTGCTGACGTCGTGCGGACCGTTTCGCGGAGATGATCAAGGTAATTGATACG	720	
QY	721	GACAAGAGATGTTCTCAAGGCGCGCGGCTTTGCAATGATTTTCGGGCCGAG	780	
DB	721	GACAAGAGATGTTCTCAAGGCGCGCGGCTTTGCAATGATTTTCGGGCCGAG	780	
QY	781	CGCGCCTGGCCGAGCGCGCTCCCGAGAGCCGAAGAGGACTGCTGCGCGCAT	840	
DB	781	CGCGCCTGGCCGAGCGCGCTCCCGAGAGCCGAAGAGGACTGCTGCGCGCAT	840	
QY	841	CTCGGATGATCGCGTGGCCAAAGCGCGCGCGATCCCGCGGCCACTATTCA	900	
DB	841	CTCGGATGATCGCGTGGCCAAAGCGCGCGCGATCCCGCGGCCACTATTCA	900	

DB	841	CTCGGATGATCGCGTGGCCAAAGCGCGCGCGATCCCGCGGCCACTATTCA	900	
QY	901	GACGTAACGGCGTGTCTGTGATCGACGTCGCGGCCCAACGCGTGTACGCTGAT	960	
DB	901	GACGTAACGGCGTGTCTGTGATCGACGTCGCGGCCCAACGCGTGTACGCTGAT	960	
QY	961	GATTCGAACCGCAAAAGAGGACCAAGGCGACGCGCGCGCTGCGGTGCGGAA	1020	
DB	961	GATTCGAACCGCAAAAGAGGACCAAGGCGACGCGCGCGCTGCGGTGCGGAA	1020	
QY	1021	AGCGCGCGCGCGCGAGTAG	1041	
DB	1021	AGCGCGCGCGCGCGAGTAG	1041	

RESULT 2  
AX189644  
LOCUS AX189644 1041 bp DNA linear PAT 08-AUG-2001  
DEFINITION Sequence 1 from Patent WO0148175.  
ACCESSION AX189644  
VERSION AX189644.1 GI:15143034  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
1. .1041  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
/note="Description of Unknown Organism: Obtained from an environmental sample"  
1. .1041  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC50776.1"  
/db\_xref="GI:15143035"  
/translation="MSEPMTRYKGAAYVQAPVFLDLDRIVEKAIGLIEQAAKQDVRLL  
AFPEWTIPGYFWIWLGAAPWGMRFVORYFENSLVRGSKQWQALADARRHGMHVAG  
YSERAGSLYMGQALFGPDGLIARRLKPTHAERTVFEGDGSILAVIDTAIGRLG  
ALCCWEHIQPLSKYAMYADEQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGG  
CYVLASCATVSPSEMIKVLVDTDPDKEMFLKAGGFAMI FPGDGRALAEPLPETEGLLV  
ADIDLGMIALAKAAADPAGHYSRPDVTIRLLDRRAQRVVTLDAFEPQNEKGDAPA  
LRVAESAAAAQ"

ORIGIN

Query Match	100.0%;	Score 1041;	DB 6;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.1e-124;		
Matches 1041;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGGCGCGCGGTGTTCTC	60	
DB	1	ATGTCGAGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGGCGCGCGGTGTTCTC	60	
QY	61	GATCTGACCGGACAGTGCAGAAAGCGATCGGCTGATCGAGCAGGCGGCCAAGCAGAG	120	
DB	61	GATCTGACCGGACAGTGCAGAAAGCGATCGGCTGATCGAGCAGGCGGCCAAGCAGAG	120	
QY	121	GTGCGCTGATCGCATTCACAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG	180	
DB	121	GTGCGCTGATCGCATTCACAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG	180	
QY	181	GGCGCGCGGCTTGGGGCATGCGCTTCGCCAGCGCTATTTCCAGAAATTCGCTCGT	240	
DB	181	GGCGCGCGGCTTGGGGCATGCGCTTCGCCAGCGCTATTTCCAGAAATTCGCTCGT	240	
QY	241	GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGGCGCGCGCGCATGATGTCGTG	300	

|||||  
Db 241 GGCAGCAAGCAGTGGCAGAGCCCTGGCCGAGTGGGCGCCGCCAGCGCATGTCATGTCGTG 300  
QY 301 GCCGGCTATAGCAGCGCGCGCGCGGAGCCTCTATATGCGGCAGGCGCATCTTCCGCCCC 360  
Db 301 GCCGGCTATAGCAGCGCGCGCGCGGAGCCTCTATATGCGGCAGGCGCATCTTCCGCCCC 360  
QY 361 GATGGCGATCTGATTCGCCCGCGCGCGCAAGCTCAAGCCTACCACATGGCGAGCGCACCGTG 420  
Db 361 GATGGCGATCTGATTCGCCCGCGCGCGCAAGCTCAAGCCTACCACATGGCGAGCGCACCGTG 420  
QY 421 TTCGGCAGGAGAGCGGACGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGC 480  
Db 421 TTCGGCAGGAGAGCGGACGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGC 480  
QY 481 GCGCTCTGTTGCTGGAGACACATCCAGCCATTGTCGAATAACGCCATGTACGCCCGCGAC 540  
Db 481 GCGCTCTGTTGCTGGAGACACATCCAGCCATTGTCGAATAACGCCATGTACGCCCGCGAC 540  
QY 541 GAACAGGTCCACGTCGCGTGTGCGCCGAGCTTCAAGCTCTATCGCGGCAATGGCCTATGCG 600  
Db 541 GAACAGGTCCACGTCGCGTGTGCGCCGAGCTTCAAGCTCTATCGCGGCAATGGCCTATGCG 600  
QY 601 CTCGGACCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660  
Db 601 CTCGGACCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660  
QY 661 GTGCTGGCGTGTGCGCGGACCGCTTTCGCCGAGATGATCAAGGTATGTTGGATACGCC 720  
Db 661 GTGCTGGCGTGTGCGCGGACCGCTTTCGCCGAGATGATCAAGGTATGTTGGATACGCC 720  
QY 721 GACAAGGAGATGTTCTTCAAGGCCGCGCGGTTTGCCATGATTTTCGGGCCCGACGGC 780  
Db 721 GACAAGGAGATGTTCTTCAAGGCCGCGCGGTTTGCCATGATTTTCGGGCCCGACGGC 780  
QY 781 CGCGCCCTGGCGGAGCCGCTCCCGGAGACCGAAGGAGTGTGTCGCCGATATCGAC 840  
Db 781 CGCGCCCTGGCGGAGCCGCTCCCGGAGACCGAAGGAGTGTGTCGCCGATATCGAC 840  
QY 841 CTCGGCATGATCGCTGTGCGCAAGCGCGCGCGATCCGCGGCGCACTATTCACGGCCC 900  
Db 841 CTCGGCATGATCGCTGTGCGCAAGCGCGCGCGATCCGCGGCGCACTATTCACGGCCC 900  
QY 901 GACGTAACCGGCTGCTGTGATGACGATCCGCGCCCAACGCGTGTCAAGCTTGATGCC 960  
Db 901 GACGTAACCGGCTGCTGTGATGACGATCCGCGCCCAACGCGTGTGTATGATGCC 960  
QY 961 GCATTGCAACCGCAAAACGAGGACAAAGGGCGACGGCGCGCTGCGGTGTGGCGAA 1020  
Db 961 GCATTGCAACCGCAAAACGAGGACAAAGGGCGACGGCGCGCTGCGGTGTGGCGAA 1020  
QY 1021 AGCGCGCGCGCGCGCAGTAG 1041  
Db 1021 AGCGCGCGCGCGCGCAGTAG 1041  
RESULT 3  
AY487438 1026 bp DNA linear ENV 05-APR-2004  
LOCUS AY487438 uncultured organism clone 2A14 nitrilase (BD7266) gene, complete cds.  
ACCESSION AY487438 GI:40890080  
VERSION AY487438.1  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1026)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis

JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1026)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1. 1026  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A14"  
/environmental\_sample  
1. 1026  
/locus\_tag="BD7266"  
/BC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97385.1"  
/db\_xref="GI:40890081"  
/translation="MISPTQYRAAVQAPSLDIDRTVEKTIATIEQAEDVRLI  
AFPETWIPGYPLWILGSPAMGRFVQRYFENSLVGSKOMNAIADARHRMTVVG  
FSERAGSLYMGQAIKPEGELIARRLKPTHAERTVFEEDGSHLAYETGVGRIG  
ALCCWEHIQPLSKYMAAYANEQVHVASWPCFSLYRGMAVALPEVNTAASQVAVEGG  
CYVLASCLVTPLEILKVLIDTPDKPLLLAGGFSMIFGPDGRALAQPLPETEEGLVT  
AEIDLGAIALAKAADPAGHYARPDVTRLNLNPRPARVEALGPRFEVQSEQAEPPT  
QPAEAD"  
ORIGIN  
Query Match 65.5%; Score 682; DB 3; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 4e-78;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;  
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTTCCTC 60  
Db 1 ATGTTAAGTCCCGTAGCAGATATCGCGCGCGGTGACAGCGCGCGCATCTTTCTC 60  
QY 61 GATTCGACCGCACAGTCGAGAAACGATCGGCTGATCGACGAGCGGCCAAGCAGAC 120  
Db 61 GATTCGACCGCACCGTCGAGAAACGATCGGATCATCGACGAGCGGCCAAGCAGAT 120  
QY 121 GTGCGCTGATCGCATTCGCCAGACTTGATTCGCCGGCTATCCCTTTTGATATGCTG 180  
Db 121 GTGCGCTGATCGCGTTTCCGGAACCTGATTCCCGGCTATCCGCTGTGATCTGGCTC 180  
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAGAATTGCTGTGCGC 240  
Db 181 GCGTCGCGCGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAGAATTGCTGTGCGC 240  
QY 241 GGCAGCAACAGTGGCAGGCGCTGCGCGGATGCGGCCCGCCAGCGCATGATGTCGTG 300  
Db 241 GGCAGCAACAGTGAACGCGATGCGCGGATGCGCGCGCGCCACCGCATGACCGTGTG 300  
QY 301 GCCGGCTATAGCAGCGCGCGCGGAGCCTCTATATGCGGCAGCGCATCTTCCGCCCC 360  
Db 301 GTCGGCTTCAAGCAGCGCGCGCGGAGGACCTCTACATGGGCCAGCGATCTTCCGCCCC 360  
QY 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCATGCGGAGCGCACCGTG 420  
Db 361 GAAGCGAGCTCATCGCGCGCGCGCGGAAGCTCAAGCCGACACACGCCGAGCAACGGTG 420  
QY 421 TTCGGCAGGAGAGCGGACGCCATCTCGCGGTGCACGATACGCCATCGGGCGCCTCGGC 480  
Db 421 TTCGGCAGGAGGAGCGGACGCCATCTGCGCGTTTACGAGCGGGCGTGTGTCATCGGC 480  
QY 481 GCGCTCTGTTGCTGGAGACATCCAGCCATTGTCGAATAACGCCATGTACGCCCGCGAC 540  
|||



Db	Accession	Version	Source	Organism	Reference	Title	Journal	Authors
Db	481	GCCCTCTGCTGCTGGAGCACATCCAGCCCGCTCTCGAATAACCGCATGTATGCGGCCAAC	540					
QY	541	GAAACAGGTCACGTCGCGTCTGGCGGAGCTTCAAGCTCTATTCGGCGATGGCCTATGCG	600					
Db	541	GAAACAGGTCATGTGGCCTCGTGGCGGCTGCTTACGCTTTATTCGGCGCATGGCCTATGCG	600					
QY	601	CTCGGACCCGGAGGTCAATACCGCGGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC	660					
Db	601	CTCGGGCCGGAGGTGAACACCGCGGAGCCAGGTCTACGCGGTGAGGGCGGCTGTAC	660					
QY	661	GTGCTGGCGTCTGTCGCGGACCGTTTTCGCGGAGATGATCAAGTATTTGTGTGATACGCCC	720					
Db	661	GTGCTGGCGTCTGTCGTCGTGTGACACCCGAGATCCTGAAGTGTGTGATTCACACGCCC	720					
QY	721	GACAAGAGATGTTCTCAAGCGCGGCGCGGCTTTTGCCATGATTTTCGGGCGCGACGAC	780					
Db	721	GACAAGAGACCGTTGTGTCTGTCGCGGCGGCGGCTTCTGATGATCTTCGCGCGCGACGAC	780					
QY	781	CGCGCCCTGGCGCGGCGGCTCCCGGAGACCGAAGAGGAGTGTGTGTCGCGCATATCGAC	840					
Db	781	CGCGCGCTCGCGCGGCGGCTCCCGGAGACCGAAGAGGAGTGTGTGTCGCGCATATCGAC	840					
QY	841	CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCGCATCCGCGCGGCCACTATTACGCGCC	900					
Db	841	CTCGCGCGCATCGCGCTCGCCAAAGCGCGCGCGCATCCGCGCGGCCACTATTACGCGCC	900					
QY	901	GACGTAAACGCGGCTGCTGTGATCGACGTCCGCGCCCAACGCGTGTGTACGCTGTATGCC	960					
Db	901	GACGTAAACGCGGTTGTGTGATCAACCGCGCGCGCGCGCGTGTGTACGCTGTATGCC	960					
QY	961	GCATTCCGAACCGCAAAACAGAGCAAGGGCGACCGCGCGCGCGTGTGTGTGCGCGG	1018					
Db	961	CGCTTCGAGTGTGTGACAGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	1018					
RESULT 4	AY487543	1062 bp	DNA	linear	ENV 05-Apr-2004			
LOCUS	AY487543							
DEFINITION	Uncultured organism clone 2A20 nitrilase (BD7753) gene, complete cds.							
ACCESSION	AY487543							
VERSION	AY487543.1	GI:40890290						
KEYWORDS	ENV.							
SOURCE	uncultured organism							
ORGANISM	uncultured organism							
REFERENCE	1 (bases 1 to 1062)							
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.							
TITLE	Exploring nitrilase sequence space for enantioselective catalysis							
JOURNAL	Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)							
REFERENCE	2 (bases 1 to 1062)							
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.							
TITLE	Direct Submission							
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955							
REFERENCE	Directors Place, San Diego, CA 92121, USA							
FEATURES	Location/Qualifiers							
source	1.1062							
	/organism="uncultured organism"							
	/mol_type="genomic DNA"							
	/db_xref="taxon:155900"							
	/clone="2A20"							
	/environmental_sample							
	1.1062							
	/locus_tag="BD7753"							
	1.1062							
	/locus_tag="BD7753"							
	/EC_number="3.5.5.7"							
CDS								
gene								

	/codon_start=1	
	/evidence=experimental	
	/product="nitrilase"	
	/protein_id="AAR97490.1"	
	/db_xref="GI:40890291"	
	/translation="MGVAHPKYKVAAVQAAAPFLDLDASVEKAVRFIDEAGAGARLLAFPEWTIPGYPMWIMLGAPAMAIMRGFSRYFDNSLSYDSPQAEKLRAAKRNKVYVIGLISERDGSLYIAQMIIIGPDGETIARKRLKPTHAERTVFEGEDGSHLVHELVDGLGALCWEHLQPLSKYAMIYAONEOVHVAWFPSLYDPFAHALGAEVNNAASKIYAVFGSCFVIAPCATVSQAMIDELCDTPEKHQFLHAGGFAYIXPDAFLAFLPPDEKGLYADIDLGMISVAKAADPAIGHYARPDTVRLLFNRRPGYRVETMALPIDAETKEAPPKPEPKAPNVAFPAPVQAEE"	
ORIGIN		
Query Match	46.7%; Score 486; DB 3; Length 1062;	
Best Local Similarity	67.8%; Pred. No. 4.5e-53;	
Matches 697; Conservative	0; Mismatches 325; Indels 6; Gaps 1;	
OY	17 CGAAGTATCGCGCGCGCGGTGCAGGCCGGCGCGGTCTTCGATCTGCACCGCACAG	76
DB	17 CGAAATACAAGTGCGCCGCGTGCAGGACGCGCGCTTTCTCGACTGCACGCTCGG	76
OY	77 TCGAGAAAGCGATCGCGCTGATCGAGCAGCGCGCAAGACGAGCGTGCGCTGATCGCAT	136
DB	77 TCGAAAAGCGCGTCCGTTTCATCGACGAAGCGCGCGCGCGCGCTCATCGCT	136
OY	137 TCCCAGAGACTTGATTCCTCCGCGCTATCCCTTTTGATATAGCTGGCGCGCGCTTGGG	196
DB	137 TTCGAGAGACCTGATACCCTGTTACCCCTGTGATCTGGCTAGCGCGCGCGCTGGG	196
OY	197 GCAT-----GCGCTTCGTCCAGCGCTATTTCAGAATTCGCTCGTGCGCGGACGAAGC	250
DB	197 CTATCATGCGCGGCTTCGTCTCGCGCTATTTCGACAACTCGCTCAGCTACGACAGCCCCG	256
OY	251 AGTGACAGGCCCTGGCGGATGCGGCCCGCGCGCAAGCATGCATGTCGTGCGCGGTATA	310
DB	257 AGGCCGAGAGACTCCGCGCGCGCGCGCGCAAGCGCAACAGATGGTGTGCTCGGCTCT	316
OY	311 GCGAGCGCGCGCGCGGACGCTCTATATGGGCCAGCGCATCTTCGCCCCGATGGCGCATC	370
DB	317 CCGAGCGCGCGCGCGGACGCTTTACATCGCGCAATGATCATCGGCCCGGACGCGCGAAA	376
OY	371 TGATCGCGCGCGCGCGCAAGCTCAAAGCTTAACCATGCGGAGCGCACCGTGTTGCGCGAG	430
DB	377 CCATCGCCAAGCGCGCGCAAGCTCAAGCCGACCCAGCGGAGCGGACCGTGTTGCGCGAAG	436
OY	431 GAGACGGCAGCCATCTCGCGGTGACGATACCGCATCGGCGCCTCGGCGCGCTTGTT	490
DB	437 GCGACGGCTCGCATCTTGCGGTGACGAGCTTGATGTGGCCGCGCTCGGCGCGCTTGCT	496
OY	491 GCTGGAGCATCCAGCCATTGTGAAATACGCCATGTACGCCGCCGACGAACAGTCC	550
DB	497 GCTGGAACACTGACAGCCGCTGTCCAATACGCCATGTATGCGCAGAACGAACAGTGC	556
OY	551 ACCTCGCGTGTGAGCGCGAGCTTCAGCTCTATCGCGCATGCGCTATGCGCTCGACCGG	610
DB	557 ATGTGCGGCGCTGGCGGAGCTTTTCGCTTAACGATCCGTTGCGCACGCGCTCGGCGCGG	616
OY	611 AGGTCAATACCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTACGTGCTGGCGT	670
DB	617 AAGTGAACATGCGCGGAGCAAAATCTATGGGTCGAGGGCTCGTGTTGTCATCGCGC	676
OY	671 CGTGCGGACCGTTTTCCCGGAGATGATCAAGTATTGGTGATACGCCCGACGAAGAGA	730
DB	677 CGTGCGGACCGTTTTCCGAGGCGATGATCGACGAACCTCGCATACGCCCGGAGAAAGCATC	736
OY	731 TGTTCTCAAGCGCGCGCGGCTTTGCGATGATTTTTCGGGCCCGACGCGCGCGCTGG	790
DB	737 AGTTCCTGATGCGCGCGCGGCTTTGCGGTGATTTACGCGCCCGACGCGCGCGCTCG	796
OY	791 CCGAGCCGCTCCCGGAGACCGAAGAGGACTGTGTCGCCGATATCGACCTTCGCGCATGA	850
DB	797 CGCGCGCGCTCCCGCGGACCGAAGGAGGCTTGCTCTACGCCGACATGATCTCGGAGTGA	856

QY	851	TCGCGTTGGCCCAAGGCGCGCGCGATCCGGCGGGCCACTATTTCACGGCCCGACGTAACGC	910
Db	857	TTTCGGTTGCCAAAAGCGGCACCGCATCCGGCCGGGCATTATGACGCCCCGACGTCACCC	916
QY	911	GGCTGCTGCTGATCGACGTCGCGGCCCAACGCGTCGTCACGCTTGATGCCGATTGCAAC	970
Db	917	GGCTTCTGTTCAACAATCGGCTTGGGTATCGGGTCGAGACCATGGCGTTGCCGATGATG	976
QY	971	CGCAAAACGAGACAAAGGCGCAGCGCCCGCTGCGCGTGGTGGCGGAAGCGCGCCG	1030
Db	977	CGAGACCAAGCGCGAAGCACCGGCTTAAGCCGGAACCCAGGACCAAGCAAGCAAGCGTGGCGCGT	1036
QY	1031	CCGCGCAG	1038
Db	1037	TCGCGCCG	1044
RESULT 5	AY487496	1065 bp	DNA linear ENV 05-APR-2004
LOCUS	AY487496		
DEFINITION	Uncultured organism clone 2A19 nitrilase (BD5217) gene, complete cds.		
ACCESSION	AY487496	GI:40890196	
VERSION	AY487496.1		
KEYWORDS	ENV.		
SOURCE	uncultured organism		
ORGANISM	uncultured organism		
REFERENCE	1 (bases 1 to 1065)		
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		
TITLE	Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)		
JOURNAL	2 (bases 1 to 1065)		
REFERENCE	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		
AUTHORS	Direct Submission		
TITLE	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955		
JOURNAL	Directors Place, San Diego, CA 92121, USA		
FEATURES	Location/Qualifiers		
source	1. 1065		
	/organism="uncultured organism"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:155900"		
	/clone="2A19"		
	/environmental_sample		
gene	1. 1065		
	/locus_tag="BD5217"		
CDS	1. 1065		
	/locus_tag="BD5217"		
	/EC_number="3.5.5.7"		
	/codon_start=1		
	/evidence=experimental		
	/product="nitrilase"		
	/protein_id="AAR97443.1"		
	/db_xref="GI:40890197"		
	/translation="MATVHPKFKVAQAAVQAPAFLLDASVEKAVRLIDEAGAAGARLI		
	AFPEFTIPGYPWWIWLGAAPAMIRGFVSRYEDNSLQYGTPEADRLRAAKRMFVA		
	LGLSERDGSLSYIAQWIIIGPDGETVATRKLLKPTHAERTVFGEGDGHILAVHIELDIGR		
	VGALCWEHLOPLSKYAMYAONEQVHIAWPSFSLYDPFAHALGAENVNAASKIYAVE		
	GSCFVIAPCATVQAMIDELCDSPKHLQPLHVGSGFAVITYGPDGAPLAPLPDQEG		
	LYADIDLGMISVAKAAADPAGHYARPDVTRLFLNNRPGRNVETLLPVDQEAAGAGG		
	KPAKSPSVAAFTLTQAAAE"		
ORIGIN			
Query Match	43.9%;	Score 457.2;	DB 3; Length 1065;
Best Local Similarity	67.2%;	Pred. No. 2.2e-49;	
Matches	664;	Conservative	0; Mismatches 318; Indels 6; Gaps 1;

QY		1 ATGTCGGAGCCCATGACGAAGTATCGCCGCCGCCGTGCAGGCCCGCCCGGTGTTCTC	60
Db		1 ATGGCGACAGTTCATCCGAAATTTAAGTAGCCGCCGCTCCAGCGCGCCCCCGGCTTTCTC	60
QY		61 GATCTCGAACCACAAGTCGAGAAAACGATCGGCTTGATCGACGAGCGGCCCAAGCAGAC	120
Db		61 GACTCGACGCGTCGGTGGAAAAAGCGGTGCGCTGATTCATGAAGCCGCGCGCTGCTG	120
QY		121 GTGCGCTGATTCGCATTTCCAGAGAATTGGATTCCCGGTATCCCTTTGGATATGCGTG	180
Db		121 GCCCGGCTCATTCGCGTTTCCAGAGACTTTTATCCCCGGTTATCCGTGTGATCTGCTC	180
QY		181 GCGCGCGCGCTTGGGGCAT-----GCGCTTCGTCCAGCGCTATTTCGAGAATTCGCTC	234
Db		181 GGTGCTCCGCGCTGGCGCATATCGCGCGGCTTGTCTCCGCTATTTCGACAATCTGCTG	240
QY		235 GTGCGCGCGCAAGCAGTCGACGACCCTGCGGATGCGGCCCGCCCAAGCATGTCAT	294
Db		241 CAGTACGGCACCCCGGAAGCCGACCGGCTGCGGCGAGCGCCCAAACGCAAAATGTTC	300
QY		295 GTCGTGCGCGGCTATAGCGAGCGCGCGCGCGCGCAGCCTCTATATGGGCCAGCATCTTC	354
Db		301 GTCGCGCTCGCACTGTACAGAGCGCGACGCGCGGCAGTCTCTACATCGCCCAATGGAATTATC	360
QY		355 GGGCCCGCATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCCTAACCCATGCGAGCGC	414
Db		361 GGACCCGACGGCGAGACGCTCGCAACGCGCGCGCAAGCTCAAGCCTACTCACGCGCGAGCGG	420
QY		415 ACCGTGTTGCGCGAGGAGACGCGAGCCATCTCGCGGTGACGATACCGGCATCGGCGC	474
Db		421 ACGGTGTTGCGCGAAGCGATGGCTCGACCTTGGCGGTCCAAGACTTGATATCGGCGCG	480
QY		475 CTCGCGCGCTCTGTTGCTGGAGCACATCCAGCATTGTGAATAACGCCATGATGCGC	534
Db		481 GTCGGTGGCTGTGCTGTGGAGACCTGCAGCCACTGTGGAAGTACGCGATGATGCG	540
QY		535 GCCGACGAACAGGTCCACGTGCGCTGTGCGCGAGCTTCAGCCTCTATCGCGGCATGCGC	594
Db		541 CAGAACGAGCAAGTTCATATCGCGCGGTGCGCGAGCTTTTCGCTTACGATCCGTTTCGG	600
QY		595 TATGCGCTCGAACCGGAGGTCAATACCGCCGCAAGCCGATCTACGCGTTCGAGGCGGC	654
Db		601 CATGCGCTTGGCGCGGAGGTCAACAACGCGCGCAGCAGATCTACGCGTTCGAGGCTCA	660
QY		655 TGCTACGCTGCGCGTGTGTCGTCGCGACCGTTTCGCGGAGATGATCAAGTATTCGTGAT	714
Db		661 TGCTTTGTGATTCGCCCATGCGCGACCGTTTCCAGGCGATGATCGACGAATTCGTGAC	720
QY		715 ACGCCCGCAAGAGATGTTCTCTCAAGGCCCGCGCGGTTTTGCCATGATTTTCGGGCC	774
Db		721 TCGCCCGAGAAAGCATCAGTTCCTGCAACGTGCGCGCGGTTTTCCGCTGATCTATGTTCCC	780
QY		775 GACGCGCGCGCCCTGCGCGGACGCTCCCGGAGACCGAAGGGAAGGACTGCTGTCGCGCAT	834
Db		781 GACGCGCGCCACTCGCCAAAGCACTGCGCGCCGATCAGAGAGGCTCTCTTTACGCGGAT	840
QY		835 ATCGACCTCGGCATGATCGCGTTGGCCAAGGCGCGCGCGATCCGCGGCGCACTATTCA	894
Db		841 ATCGACCTCGGCATGATTTCCGTCGCGAAGCGCGCGCGATCCGCGTGCACATTACGCG	900
QY		895 CGGCGCGACGTAAAGCGCGCTGCTGTGATCGACGTCGCGCCCAACGCGTCTGTCAAGCTT	954
Db		901 CGCCCGACGTGACCCGCTGTGTTGTTCAACAATCGTCTGGGAACCGGGTGAACACCTC	960
QY		955 GATGCGCGATTCGAACCCGCAAAACGAGG 982	
Db		961 GCGCTGCTGCTGACCAAGCGACAGG 988	

RESULT 6  
AY487522      1077 bp    DNA    linear    ENV 05-APR-2004

LOCUS

DEFINITION Uncultured organism clone 2A17 nitrilase (BD5263) gene, complete cds.  
ACCESSION AY487522  
VERSION AY487522.1 GI:40890248  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1077)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1077)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA  
FEATURES  
source 1..1077  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A17"  
/environmental\_sample 1..1077  
/locus\_tag="BD5263"  
1..1077  
/locus\_tag="BD5263"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97469.1"  
/db\_xref="GI:40890249"  
/translation="MALTHRLKVAAVQAPAFLDVAVDKAVRLIDEAANGSSLV  
AFPETWIPGYFWILGSPAWIMRGFVSRYFDNSRQAERLRDAKRHLKTVY  
MGLSERAGGSLYIAQWITGNGETVAQRKLKPTHAERTVFEGBDLSHLAVHNLPIGR  
LGALICMEHLQPLSKYAMYNGEIEIHAAMPSESLYDPEAHAGAEVNNAASQIYAVE  
GSCFVAPCAVISQEMIDLCDPKHQLIHVGGFYIYGPDGARIGDKLAPDEGI  
VYADIDLGIMPIAKAADPAGHYARPDVTRLFFNNRPNRVETLVLPDQVRDIDARV  
EAAAPQARPATGNEDPAKPMAAE"  
ORIGIN  
Query Match 43.7%; Score 455.2; DB 3; Length 1077;  
Best Local Similarity 67.9%; Pred. No. 3.9e-49;  
Matches 653; Conservative 0; Mismatches 303; Indels 6; Gaps 1;

QY 311 GCGAGCGCGCGCGCGCAGCCTCTATATGGCCAGGCGATCTTCGCCCCGATGGCGATC 370  
DB 317 CCGAGCGCGCGCGCGGTAGCCTTTACATCGCGCAGTGATCATTTGGTCCCATGGCGAGA 376  
QY 371 TGATCGCGCGCGCGCGCAGCTCAAGCCTTACCATGCGGAGCGCACCGTGTCCGCGAGG 430  
DB 377 CCGTCGACAGCGCGCGCAGCTCAAGCCACCCATCGGAGCGCACCGTCTTCGCGCAGG 436  
QY 431 GAGACGAGCAGCCATCTCGCGGTGCAAGATACCGCCATCGGAGCGCTCGGCGCTCTGTT 490  
DB 437 GTGACGCGAGCCACCTGGCGGTACACAATCTTCCAATCGGACGGCTCGGTGCGTGTGCT 496  
QY 491 GCTGGAGCACATCCAGCCATTGTGAAATACGCCCATGACGCCCGGAGCAACAGATCC 550  
DB 497 GCTGGAGCACCTCCAGCCGCTCTCAATACGGCATGTACGCCAGAAAGAGATCC 556  
QY 551 ACGTCGCGTGTGGCGGAGCTTCAGCCTCTATCGCGGATGGCCTTATGCGCTCGACCGG 610  
DB 557 ACGTGGCGCATGGCGCGTCTTCTGCTCTACGACCCGTTTGGCACGCGCTCGCGCCG 616  
QY 611 AGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTACGTGCTGGCGT 670  
DB 617 AAGTCACAACAGCAGCAGCAGCATCTACGCGTTGAAGGTTCTGCTTGTGTCGCGC 676  
QY 671 CGTGGCGACCGTTTCGCGGAGATGATCAAGTATTGTGTGATACGCCGCAAGAGA 730  
DB 677 CATGTGCGGTGATCTCGCAGGAATGATCGATCTTATGTGCGATACCCCGACAAGCATC 736  
QY 731 TGTTCCTCAAGCGCGCGCGGTTTTCGCATGATTTTCGCGCCCGACGCGCGCCCTGG 790  
DB 737 AGCTTATTCAGTCGTCGTCGCGCTTCAACGTGATCTATGCGCCCGACGTCGCGCATCG 796  
QY 791 CCGAGCGGCTCCCGGAGACCCGAAGAGGACTGTCGTGCGCGCATATCGACCTCGGCATGA 850  
DB 797 GCGACAAGCTCGCGCCAGATCAGGAAGGATTGTCTATGCCGACATCGATCTCGGCATGA 856  
QY 851 TCGCGTTGGCCCAAGCGCGCGCGCATCCGGCGGCGCACTATTCACGCGCCGACGTAAAGC 910  
DB 857 TCCCGATCGGAAAGCTGCGCGCGATCTGCGCGCCACTATGCGGACCCGACGCTTACCC 916  
QY 911 GCGTGTCTGTGATCGACGCTCGCGCCCAACGCGTCTGTCACGCTTGATGCCGATTCGAAC 970  
DB 917 GCCTTCTGTTCAACAATCGTCCCGCAATCGGGTGAAACCTCTGTCTCCCGTTGATC 976  
QY 971 CG 972  
DB 977 AG 978  
RESULT 7  
AY487536 1017 bp DNA linear ENV 05-APR-2004  
LOCUS  
DEFINITION Uncultured organism clone 2A29 nitrilase (BD5279) gene, complete cds.  
ACCESSION AY487536  
VERSION AY487536.1 GI:40890276  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1017)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1017)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission



JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1.1017  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A29"  
/environmental\_sample  
1.1017  
/locus\_tag="BD5279"  
1.1017  
/locus\_tag="BD5279"  
/locus\_tag="BD5279"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97483.1"  
/db\_xref="GI:40890277"  
/translation="MGIHPKYRVAVQAAPAWLDLDRSIDKALIEBAANGARLI  
AFPEVPIPGYPWHIWLDSPAWALGRFVQRYFDNSLAYDSPQERLRAVRKARLTAV  
IGLSERSGSLYIAQWLVPDGETAKRKLRPTHAERTVYEGSDSLAVHDPDIG  
RLGALCCWEHLQPLSKYAMVAONEVHVAWSPSLYDPFAPALGAENVNAASRVAV  
EGSCFYALPCATVSQAMIDELCDRPKHALLHAGGFAIYGDGSSLAEKLPDQEG  
LTYADIDLGAIGVAKNADPAGHYSRPDVTRLLLNKPKYRVEHFALPGDTVAPADVD  
AAS"

ORIGIN

Query Match 43.7%; Score 454.6; DB 3; Length 1017;  
Best Local Similarity 67.5%; Pred. No. 4,7e-49;  
Matches 672; Conservative 0; Mismatches 314; Indels 9; Gaps 2;

QY 17 CGAAGTATCGCGCGCGCGCGGTGTCAGAGCGCGCGGTGTTCTCGATCTGACCGCACAG 76  
DB 17 CAAATACCGCGTTCGCCCGGTGTCAGAGCGCGCGCGCTGCTCGATCGATCGCTCGA 76  
QY 77 TCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAGCGATGCGCTGATCGCAT 136  
DB 77 TCGACAAGGCTATTGCACTGATCGAGAGCGCGCGCAAGCGCGCATGTATCGCAT 136  
QY 137 TCCAGAGACTTGATATCCCGGCTATCCCTTTGATATGCTGGCGCGCGCTTG- 195  
DB 137 TCCCGAGGCTTCATCCCCGGCTACCCCTGSCATATCTGGCTCGATCGCGCGCTGG 196  
QY 196 -----GGCATGCGCTTCTGTCAGCGCTATTTCAGAAATTCGCTCGTGGCGCGCAGCAAGC 250  
DB 197 CGATCGCGCGCGCTTCTGTCAGCGCTATTTCAGCAATTGCTCGCTTATGATAGTCCGC 256  
QY 251 AGTGCAGGCGCTGCGCGATGCGCGCGCGCGCGCATGATGCTGCGCGCTATA 310  
DB 257 AGGCCAGCGGCTCGCGCAGCGGTCCGCAAGCGCGCTGACCGCGCTGATCGGCTTT 316  
QY 311 GCGAGCGCGCGCGCGCAGCCTTATATGCGCGCAGCGCATCTTCGCGCGCATGGCATC 370  
DB 317 CGGAGCGGAGCGCGCGCAGCCTTACATCGCGCATGGCTCGTTGCGCGCAGCGGAGA 376  
QY 371 TGATCGCGCGCGCGCGCAGCTCAAGCCTACCATGCGGAGCGCACCGTGTTCGGCAGG 430  
DB 377 CCATCGCGAGCGCGCGCAAGCTCCGTCGACGCAATGCGGAGCGCACCGTCTATGGCAGG 436  
QY 431 GAGACGGCAGCCATCTCGCGGTGACGA---TACCGCATCGGGCGCTCGCGCGCTCT 487  
DB 437 GCGACGGCAGCGATCTGGCGGTCCATGACCGCGCGCATATCGGACGGCTCGCGCGCTGT 496  
QY 488 GTTGCTGGAGACATCCAGCCATTGTGAAATACGCCATGTACGCGCGCAGCAAGAG 547  
DB 497 GCTGCTGGGAACAATCTGCAACCGTTGTGAAATATGCGATGTATGCCAGAAAGAGCAGG 556  
QY 548 TCCACGTCGCGTGTGGCGCGAGCTTCAAGCCTTATCGCGCATGCGCTATGCGCTCGAC 607  
DB 557 TCCATGTGGCGTATGGCCGAGGTTTTCGCTTACGATCCCTTGGCCCGCGCTCGGCG 616  
QY 608 CGGAGGTCAATACCGCGCGCAAGCAGATCTACGCGGTGAGGCGCGCTGCTACGTCTGG 667

DB 617 CGGAGGTGAACATGCGCGCGCTCCCGGCTATGCGGTGAGAGGCTCCTGCTGCTGG 676  
QY 668 CGTCGTGCGCGACCGTTTCGCGCGGAGATGATCAAGGATTTGGTGATACGCGCGCAGAG 727  
DB 677 CGCGGTGCGCGACCGTCTCGCAGGCGCATGATGAGCTGTGCGACCGCGCGCAGCAGC 736  
QY 728 AGATGTTCTCAAGCGCGCGCGGTTTGGCCATGATTTTCGGGCGCGCAGCGCGCGCC 787  
DB 737 ACGGCTGCTCATGCGCGCGGTGGCTTTGCGCGATCTACGGGCGCGCAGGAGTTCCG 796  
QY 788 TGGCGGAGCGCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCGCATATCGACCTCGCA 847  
DB 797 TGGCGGAAAGCTCGCGCGGACGAGAGGCGCTGCTTACGCGCAGCATCGATCTCGCG 856  
QY 848 TGATCGCGTTGGCCCAAGCGCGCGCGCATCCGCGCGCCACTATTCAAGCGCGCGAGTAA 907  
DB 857 CGATCGGCGTCCGGAAGACCGCGCGCAGCGCGCATTTTCGCGCGCGCATGTCA 916  
QY 908 CGCGGCTGCTGCTGATCGACGCTCGCGCGCCCAACGCGCTGTCACGCTGATGCGCATTCG 967  
DB 917 CGCGGCTGCTGCTGAACAACAAGCCCTTACAAGCGCGGTGAGCATTTTGTGCCCCGCG 976  
QY 968 AACCGCAAAACGAGCAGCAGGCGCGCGCGCGCGCG 1002  
DB 977 ATACCGTGGCGCGCTGCGCATGTGATGCGCGCGCG 1011

RESULT 8  
AY487558 1017 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A31 nitrilase (BD5287) gene, complete  
DEFINITION cds.  
ACCESSION AY487558  
VERSION AY487558.1 GI:40890320  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
AUTHORS 1 (bases 1 to 1017)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1017)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission  
TITLE  
JOURNAL  
SUBMITTED (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1.1017  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A31"  
/environmental\_sample  
1.1017  
/locus\_tag="BD5287"  
1.1017  
/locus\_tag="BD5287"  
/locus\_tag="BD5287"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97505.1"  
/db\_xref="GI:40890321"  
/translation="MGIQHPKYRVAVQAAPAWLDLBSVSKSIALIEBAAKAKALI  
AFPEAFIPGYPMWILDSPAWALGRFVQRYFDNSLSDSPAERLRLAVKXGMATVAV



LGLSERDGSLSLYLAQWLI GPDGETIAKRRLRPTHAERTVYEGDSDLAVHDPGIG  
RLGALCWEHLQPLSKYAMYAQNQEVHVAWPSFSLYDPFAPALGWEVNNASRVYAV  
EGSCFVLAPCATVSOAMIDELCDRDKHALLHVGGHAAIFGPDGSAIADKLPSDOEG  
LIFADIDLGAIGIAKNAADPAGHYSRPDVTRLLINKPKSRVEHFALPLDTLAGEEID  
AAAS"

ORIGIN

Query Match		43.6%;	Score 453.6;	DB 3;	Length 1017;
Best Local Similarity		67.5%;	Pred. No. 6.3e-49;		
Matches 671;		Conservative 0;	Mismatches 314;	Indels 9;	Gaps 2;
QY	17	CGAAGTATCGCGCGCGCGGTGCAGAGCGCGCGCGGTTCCTCGATCTGCACCGCACAG	76		
Db	17	CGAATATATCGCGTCCGCGGTGTGCAGAGCGCGCACCGCGCTGGCTCGACCTCGAGCGGCTCGG	76		
QY	77	TCGAGAAAGCGATCGCGCTTGATCGAGCAGCGGCCCAAGCAGACGCTGCGCTGATCGCAT	136		
Db	77	TCAGCAAGAGCATCGCGCTGATAGAGAGCGCGCCCAAGGGCGCCCAAGCTGATCGCGT	136		
QY	137	TCCAGAGACTTGGAATCCCGGCTATCCCTTTTGGATATGGCTGGCGCGCGGCTTGG-	195		
Db	137	TCCCGAGGCGCTTCAATCCCGGCTATCCCTGGTACATCTGGCTGACTTGCAGCGGCTGGG	196		
QY	196	-----GGCATCGCGCTTCTGTCAGCGCTATTTGAGAATTCGCTCGTGCAGCGGCGCAAGC	250		
Db	197	CGATCGCGCGCGGCTTCTGTCAGCGCTATTTGACAATTCGCTCAGCTATGACAGCGCGC	256		
QY	251	AGTGGCAGGCGCTGCGCGGATCGGCGCGCGCGCGCGCATGCTGCTGGCGGCTATA	310		
Db	257	AGCGGAGCGGCTGAGGCTCGCAGTGAAGAAGCGCGCATGACCGCAGTGTCTCGGCTGT	316		
QY	311	GGAGCGCGCGCGCGCGCAGCCTCTATATGGGCGAGCGCATCTTGGCGCGCGATGCGCATC	370		
Db	317	CGAGCGCGCGCGCGCGCAGCCTCTATCTCGCGCAATGTTGATCGGACCGCGCGGAGAGA	376		
QY	371	TGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCTGCGGAGCGCACCGTGTTCGGCGAGG	430		
Db	377	CCATCGCAAGCGCGCGCAAGCTCGCGCGCGCACCCATGCGGAGCGCACCGTCTACGGCGAGG	436		
QY	431	GAGACGGCGAGCCATCTCGCGGTGCACGA---TACCGCCATCGGCGCGCTCGGCGGCTCT	487		
Db	437	GCGACGGCGAGGACCTTGCCTGTGATGACCGCGCGCGCATCGCGCGCTCGTGCCTGT	496		
QY	488	GTTGCTGGGAGCATCTCAGCCATTGTGAAATACGCCATGTACCGCGCGCGAGCAACAGG	547		
Db	497	GCTGCTGGGAGCATCTGCAGCGCGCTGTGAAATACGCCATGTACCGCGCGCGAGCAGG	556		
QY	548	TCCACGTGCGGCTGTGGCGCGAGCTTCAAGCTTATTCGCGCGCATGCGCTATGCGCTCGAC	607		
Db	557	TGCATGTGCGGCGCTGTGGCGAGCTTCTGCTGTACGATCCGTTCCGCGCGCGCTCGGCT	616		
QY	608	CGAGGTCAATACCGCGCGAAGCCAGATCTACGCGTGCAGGCGCGCTGTACGTGTGG	667		
Db	617	GCGAGGTCAACATGCGGCGCTGTGCTATGCGGCTGTGAGGCGCTCTGCTTGTGTGG	676		
QY	668	CGTCTGCGCGAGCCGTTTCGCGCGAGATGATCAAGGTATGTTGATGATACGCCCGACAAGG	727		
Db	677	CGCCTGCGCGACCGTCTCGAGCGATGATGACGAGCTTGCAGCCGCGACCGAGACAAGC	736		
QY	728	AGATGTTCTCAAGCGCGCGCGGTTTGGCATGATTTTCGGCGCGCGAGCGCGCGCC	787		
Db	737	ATGCGTGTGTCATGTTGGCGCGCGCATGCGCGATCTTCCGCCCGAGCGGAGCGGA	796		
QY	788	TGCGCGAGCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCGCGATATGACCTCGGA	847		
Db	797	TGCGCGACAAGCTTCCGTCCGACCGAGAGGCGCTCTGTTCGCGACATCGATCTCGCG	856		
QY	848	TGATCGGTTGGCCAAGCGCGCGCGATCCGGCGGCGCACTATTCACGGCGCGAGCTAA	907		
Db	857	CGATCGGATTCGCAAGAATGCGCGTGTATCCGGCGGCACTATTCGGCGCGGAGCTGA	916		
QY	908	CGCGGCTGTGCTGATGACGTCCGGCGCAAGCGGTGTCAAGCTTGATGCCGATTCG	967		

Db	917	CGCGGCTGTGCTCAACACAGAAGCCCTCGAAGCGCGCTCGAGCATTCCGCTGCCGCTCG	976
QY	968	AACCGCAAAACGAGACAGGCGCGAGCGCGCCCGC	1001
Db	977	ACAGCGTCCGGCGCGAGAGATCGACGCGCGCC	1010

RESULT 9  
AY487475  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

AY487475 1017 bp DNA linear ENV 05-APR-2004  
Uncultured organism clone 2A28 nitrilase (BD7045) gene, complete  
cds.  
AY487475 GI:40890154  
ENV.  
uncultured organism  
uncultured organism  
unclassified; environmental samples.  
1 (bases 1 to 1017)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1017)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
Location/Qualifiers  
1..1017  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A28"  
/environmental\_sample  
1..1017  
/locus\_tag="BD7045"  
1..1017  
/locus\_tag="BD7045"  
/locus\_tag="BD7045"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97422.1"  
/db\_xref="GI:40890155"  
/translation="MGIEHPKYKVAVQAAPAWLDLSDIKTGLIEBAKKGAKLI  
AFPEAFIPGYPMHWLSDPAMWIGRGFVGRYEDNSLAYDSPQAERLRQAVRKAULTAV  
IGLSERDGSLSLYLAQWLI GPDGETIAKRRLRPTHAERTVYEGDSDLAVERADIG  
RLGALCWEHLQPLSKFAMYAQNQEVHVAWPSFSLYDPFAPALGAEVNNAASRIYAV  
EGSCFVLAPCATVSOAMIDELCDRDKHALLHAGGFAAIYGPDSQIGDKLPPEQEG  
LIIAIEDLGAIGIAKNAADPAGHYSRPDVTRLLINKPKRVEQFALPVTIPEPVDVA  
AAAS"

ORIGIN

Query Match		43.3%;	Score 450.4;	DB 3;	Length 1017;
Best Local Similarity		67.3%;	Pred. No. 1.6e-48;		
Matches 669;		Conservative 0;	Mismatches 316;	Indels 9;	Gaps 2;
QY	17	CGAAGTATCGCGCGCGCGGTGCAGGCGCGCGCGGTTCCTCGATCTGCACCGCACAG	76		
Db	17	CGAATATCAAGGTCCGCGGTGTGCAAGCTGCGCGCGCGCTGCTCGACCTCGAGCGGCTGA	76		
QY	77	TCGAGAAAGCGATCGCGCTTGATCGAGCAGCGCGGCCAAGCAGCTGCGCTGATTCGAT	136		
Db	77	TCGACAAAGCATCGCGCTTGATCGAGGAGCGCGCGCAAGAGCGCAAGCTGATTCGCTT	136		
QY	137	TCCAGAGACTTGATTCGCGGCTATCCCTTTTGGATATGGCTGGCGCGCGGCTTGGG	196		

Db 137 TCCCCGAAGCCTTCATTCCCGGCTACCCCTTGGCATCTGCTCGACTACACCCGCTGGG 196  
QY 197 GCATGCGC-----TTCGTCAGCGCTATTTCGGAATTCGCTCGTGGCGGACGAAGC 250  
Db 197 CGATCGGCGCGGTTTCGTGACGCGCTATTTCGACATTCGCTCGCCTACGACAGCCAC 256  
QY 251 AGTGCAGAGCCCTGCGCGGATGCGGCGCGCCGACGCGCATGCTGCTGCGCGCTATA 310  
Db 257 AGCGGAAAGGCTGCGACAGGCGCGTGCAGAGGCCAAGCTCACCGCCGTGATCGGCTGT 316  
QY 311 GCGAGCGCGCGGCGGACGCTCTATATGGCCAGCGCATCTTGGCGCCGATGGCGATC 370  
Db 317 CCGAGCGCGAGCGCGGAGCCTCTATCTCGCGAGTGGCTGATCGGCGCCGACGCTGAGA 376  
QY 371 TGATCGCGCGCGCGGCAAGCTCAAGCCTTACCATGCGAGCGCACCGTGTTCGGCAGG 430  
Db 377 CCATCGCAAAGCGCGCAAGCTGCGCGCGACCCATGCGAGCGCACCGTCTATGGCGAAG 436  
QY 431 GAGACGACGCACTCTCGCGGTGACGA--TACCGCATCGGCGCCTCGCGCGCTCT 487  
Db 437 GCGACGCGACGATCTCGCGCTCCATGAGCGGCGGACATCGCGCGGCTCGCGCGCTGT 496  
QY 488 GTTGCTGGAGACATCCAGCCATTGCGAATACGCCATGTACGCGCGCGGACGAACAGG 547  
Db 497 GCTGCTGGAGCATCTGCAAGCCGCTGTGAAATTCGCCATGTACGCCCGCAGAACGACAG 556  
QY 548 TCCACGTCGCGTCTGCGCGGAGCTTCAAGCCTTATCGCGGCGATGCGCCTATGCGCTCG 607  
Db 557 TACATGTGCGCGCTGCGCGGAGCTTCTCGCTTACGATCCCTTCGCGCTCGCTGGCG 616  
QY 608 CGGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTGCTAGTGTG 667  
Db 617 CGGAGTGAACAACGCGCGCTCCGCACTATGCGGTGGAAGGCTCCTGCTTGTGCTCG 676  
QY 668 CGTGTGCGGACGCTTTCGCGGAGATGATCAAGTATTGTGATACGCCCGACGAAGG 727  
Db 677 CACGCTGCGGACGCTCTCGAGGCCATGACGAGCTCTGCGATCGGCGGACGAAGC 736  
QY 728 AGATGTTCTCAAGCGCGCGCGGCTTTCGCAATGATTTTCGGGCGCCGACGCGCGCCC 787  
Db 737 ACGCGCTGCTGATGCGCGCGCGGCTTTCGCGCGATCTACGGGCGCCGACGCGACGCA 796  
QY 788 TGGCCGAGCGCGCTCCCGGAGACCGGAAGAGGACTGCTGTGCGCGCATGACCTCGGCA 847  
Db 797 TCGGCGACAAGCTGCGCGCGCGGAGCAGAGGCGCTGCTGATCGCGGATGATCTGGGCG 856  
QY 848 TGATCGCTTGCCCAAGCGCGCGCGCATCCGCGGCGCCACTATTACGCGCGCGACGTAA 907  
Db 857 CGATCGCGCTGCGCAAGACGCGCGCATCCGCGCGGCAATTATTGCGCGCGCGACGTCA 916  
QY 908 CGCGGCTGCTGTGATCGACGTCCGCGCAAGCGGTGTCACGCTTGAATGCGCATTCG 967  
Db 917 CGCGGCTCCTGCTCAACAGGAAGCGCAACAGCGGTGAGACAGTTCGCGCTGCCCGTGC 976  
QY 968 AACCGCAAAACGAGACAAGGGCGACGCGCGCC 1001  
Db 977 ACACGCTCGAGCCCGTCAAGCTCGCGCGGCGAGC 1010

RESULT 10  
AY487559 1005 bp DNA linear ENV 05-APR-2004  
LOCUS AY487559  
DEFINITION Uncultured organism clone 2A36 nitrilase (BD7051) gene, complete cds.  
ACCESSION AY487559  
VERSION AY487559.1 GI:40890322  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1005)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,

TITLE McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
JOURNAL Exploring nitrilase sequence space for enantioselective catalysis  
REFERENCE Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
AUTHORS 2 (bases 1 to 1005)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA  
FEATURES location/Qualifiers  
source 1..1005  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A36"  
/environmental\_sample  
1..1005  
/locus\_tag="BD7051"  
gene 1..1005  
/locus\_tag="BD7051"  
CDS 1..1005  
/locus\_tag="BD7051"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97506.1"  
/db\_xref="GI:40890323"  
/translation="MGTKHFAFKAAVVQAPEWLDLRTVDKTLALIEAAGAKLI  
APPETWIGYPWHIWGTPMAWISRGFVQRYFDNSLAYDSPAQRADAAKKNKTV  
LGLSEREGSLYISQWLIGPDETIARKRLRPTHVERTVFGDGDSHIAVHERADIG  
RLGLCCWEHIQPLTXVAMVAQNEQVHVAWPSFSMYEPFAHLGMEVNNAAASKIYAV  
EGSCFVLGACAVISQANVDEMCDTEBKRALVHAGGHAIVFGPDGRSLADKIPETQEG  
LLYADIDLGAIGVAKNADPAHYSRPDVTRLFLFNKPARRVEYFSLPDAVETPPQP  
"  
ORIGIN  
Query Match 43.2%; Score 450; DB 3; Length 1005;  
Best Local Similarity 69.4%; Pred. No. 1.8e-48;  
Matches 644; Conservative 0; Mismatches 275; Indels 9; Gaps 2;  
QY 28 GCGCGCGCGGTGCAAGCGCGCGCGGTTCCTCGATCTGCACCGCAAGTCGAGAAAGCG 87  
Db 28 GCGCGAGTGTCCAGGCGCGCGCGGATGGCTCGATCTGCACCGCAAGTCGAGAAAGCG 87  
QY 88 ATCGCGCTGATCGAGCAGCGCGCGCAAGCAGTGCCTGATTCGATTCGCCAGAGACT 147  
Db 88 ATCGCGCTGATCGAGAGCGCGCGCGCGCGCGCAAGCTCATTTGCTCCCGAAACC 147  
QY 148 TGATTCGCGGCTATCCCTTTTGATATGCTGGCGCGCGGCTTGGG-----GCATG 201  
Db 148 TGATTCGCGGCTATCCGTGGACACTCTGGGTCCGACGCGCGCGGTGGCGATCAGCCG 207  
QY 202 CGCTTCGTCAGCGCTATTTCGAGAATTCGCTCGTGGCGCGCAAGCAGTGGCAGGCC 261  
Db 208 GGCCTTCGTGACGCGCTACTTCGACAAATTCAGTGGCCTACGACAGCCCGCAGCGCC 267  
QY 262 CTGGCGGATGCGCGCGCGCGCGCGCATGCTGCTGGCGCGGCTATAGCGAGCGCG 321  
Db 268 ATCGCGGAGCGCGCGGAGAAAGAAATCAACGCTGTGCTCGGCTGTCGAGCGCGAG 327  
QY 322 GCGCGAGCCTTATATGCGCAGCGGATCTTGGCGCGCGATGCGGATGATGCGCGCG 381  
Db 328 GGTGGCAGCCTTATATCTCGCAGTGGCTGATTGGCGCGGCGAGACCATTGCCAAAG 387  
QY 382 CGCGGCAAGCTCAAGCCTTACCAATGCGGAGCGCACCGTGTTCGCGAGGAGCGAGC 441  
Db 388 CGCGGCAAGCTCGCGCGCGCGCGCGCGCATGCGCACCGTGTTCGCGAGCGGAGC 447  
QY 442 CATCTCGCGGTGACGA--TACCGCATCGGCGCGCTCGCGCGCTCTGTGCTGGAG 498  
Db 448 CACATCGCGGTGACGAGCGCTGTGACATCGGCGCGCTCGCGCGCTGTGCTGGAG 507

OY		499	CACATCCAGCCATTGTCGAATAACGCCATGTACGCCCGCCGACAACAGGTCACGTCGGC	558
Db		508	CACATCCAGCCGCTGACCAAATAACGCATGTATGCCAGAACGACAGGTGCACGTCGCC	567
OY		559	TGCTGGCCGAGCTTCACTCTATTCGGGCATGCGCCTATGCGCTCGAACCAGAGTCAAT	618
Db		568	GCCTGGCCGAGCTTCTCGATGTACGAGCCGTTGCCCCACGCGCTTCGGCTGGAAAGTCAAC	627
OY		619	ACCGCCGCAAGCCAGATCTACCGCGTCGAGGGCGGCTGCTACGTGCTGCGCGC	678
Db		628	AATGCGGCGAGAAGATCTACGCCGTCGAAGGCTCGTGTTCGTGCTCGCGCATGCGCG	687
OY		679	ACCGTTTCGCGGAGATGATCAAGGTATTGTGATACGCCCGACAAGAGATGTTCTC	738
Db		688	GTCATCTCGCAGCGATGATGTCACGAATGTGCACACCGAGGACAAGCGGGCGTGTC	747
OY		739	AAGCGCGGCGGCTTTGGCATGATTTTCGGGCCGACGCGCGCCCTGGCCGAGCCG	798
Db		748	CATGCCGCGGCGGCCACGCGCGGTGATCTTCGGGCCGACGCGCATCGCTGGCGGACAAG	807
OY		799	CTCCCGGAGACCGAAGAGGACTCTGTCGCCGATATCGACTTCGGCATGATCGCGTTG	858
Db		808	ATTCCGGAGAGCCCAAGAGGCTGCTCTATGCCGACATCGACTTCGGCGCAATTGGCGTG	867
OY		859	GCCAAGGCGGCGCCGATCCGCGGGGCCACTATTTCACGCGCCGACGTAAACGCGGCTGCTG	918
Db		868	GCCAAGAACGCGGCCGATCCGCGGGGCCACTACTCGCGCCCAGGACGTGACGCGGCTCTTG	927
OY		919	CTGGATCGACGTCCGCGCCCAACGCGTCG	946
Db		928	TTCAACACAAAGCCGCGCGCGCTGGGTCG	955

```

RESULT 11
AY487430
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS

```

```

/evidence=experimental
/product="nitriylase"
/protein_id="AAR97377.1"
/db_xref="GI:40890065"
/translation="MPTSKQFRVAVQAAVPFLDLEGALISKGISLIEAASNGAKLIA
FPETWIPGYPMWIMLDSPAWGMRRVQRYFDNSIMLGSEQAKRNQAANNKIYVMGY
SERGGSLYMGQSIINDKGETIFTRRKLPKTHVERTVFEGEGDGHLCVMDTEIGRVGA
MCCWEHLQPLSKYAMYSQDEQIHIAWSPSFLYRGAAVALGPFLNNAASQMYAAEGQC
FVLAPCATVSKEMIEMLIDDPKREPLLLEGGFMTIYGPDPGRAPLPENEGGLLYA
DIDLGMI SMAKAADPAGHYARPDVTRLLFNSAPANRVEYINPASGPTESLKDMGMQM
MEAEQQKALREMI"

```

Query Match	43.1%;	Score 448.6;	DB 3;	Length 1047;
Best Local Similarity	67.7%;	Pred. No. 2.7e-48;		
Matches 628;	Conservative 0;	Mismatches 299;	Indels 0;	Gaps 0;
QY	20	AGTATCGCGCGCGCGCGGTGCAGGCCGCGCGGTCTCTCGATCTCGACCGCAGTGC	79	
Db	17	AATTAGATCGCTGCAGTTCAGGCCGCGCGGTATTCTTGACCTGAGGCGCAATAA	76	
QY	80	AGAAAGCATCGGCTGTATCGAGCAGCGGCCAAGCAGAGCGTGCCTGATTCGATTCC	139	
Db	77	GCAAGGCAITCTCCCTCATTTAGAGAGCGCGTTCCAATGAGCCAAAGTCATTGCC	136	
QY	140	CAGAGACTTGATTCGCGCTATCCCTTTGATATGCTGGCGCGCGCTTGGGCA	199	
Db	137	CGAAACGTGATTCGCGCTACCCCTGTGATCTGCTGACTCACCGCTTGGGCA	196	
QY	200	TGCGCTTGTCCAGCGCTATTTCGAAATTCGCTCGTGGCGGACAGACAGTGGCAGG	259	
Db	197	TGCGCTTGTCCAGCGCTATTTCGAAACTCGCTCATGCTGGTAGTGAACAACCAAGC	256	
QY	260	CCCTGGCGATGCGGCGCGCGCGCACGGCATGATGCTGCGCGCTATAGCAGCGCG	319	
Db	257	GCATGAACCAAGCGCTGCCCGCAATACAGATTTACGTGTGATGGTTATAGCAACGA	316	
QY	320	CGGCGGAGCGCTTATATGGGCCAGGCGATCTTGGCGCGCATGGCATCTGATCGCGG	379	
Db	317	GTGCGGCGAGCTTACATGGGCCAATCATTATCAACGACAAGGTGAACGATTTT	376	
QY	380	CGCGCGCAAGCTCAAGCTTACCATGGCGAGCGCAACCGTGTTCGGCGAGGAGACGGCA	439	
Db	377	CCCGCGCAACTCAAGCCAACTCATGTGAGCGGTACCCTGTTGGGAGGAGACGGCA	436	
QY	440	GCCATCTCGCGGTGCACGATACCGCCATCGGGCGCTCGGCGCGCTGTGTCTGGGAGC	499	
Db	437	GCCATCTTGCCTAATGATACCGAGATTGGCGCGCTCGGCGCGATGTCTGTGGGAA	496	
QY	500	ACATCCAGCATTTGTGAAATACGCCATGTACGCCCGCGCAAGACAGTCCAGTCCGCT	559	
Db	497	ATTGCAAGCGCTCAGCAAAATATGCAATGTATTCTCAGATGAACAAATTCACATTGCT	556	
QY	560	CGTGGCGAGCTTCAACCTCTATCGCGCATGCGCTATGCGTCCGACCGAGGTCAATA	619	
Db	557	CGTGGCGAGCTTTCGTTATATCGGGGCGACCTATGCACCTCGGCCCGAACTGAACA	616	
QY	620	CCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCTGTACGTGCTGCGCTGCGCGCA	679	
Db	617	ACGCGCGCAAGCAAAATGTATGACGCCGAGGCGCAGTGTCTTGTCTTGCCTTGGCGCA	676	
QY	680	CCGTTTCGCGGAGATGATCAAGTATGTGATACGCCCGCAAGAGAGATGTTCTCA	739	
Db	677	CCGTCTCAAGAGATGATCGAAATGCTGATAGATGATCCCAAGAAAGACCGCTTCTGC	736	
QY	740	AGGCGGCGCGGTTTGGCATGATTTGGGCGCGGCGCGCGCGCTTGGCGCGAGCGCGC	799	
Db	737	TGGAAGGTGGCGGTTTCAACCATGATTTACGCGCGCGGATGGGCGACCTCTGGCTAAACCGT	796	
QY	800	TCCCGGAGACCGAAGAGGAGCTGCTGGTCCCGCATATCGACCTCGGAGTATCGCGTTGG	859	
Db	797	TGCTTGAACACGAGGAGGCTGTATATGCGCATATTGACCTGGGCGATGATTTCAATGG	856	



QY 860 CCAAGCGCGCCGATCCGGCGGCCACTATTCAGCGCCCGACGTAACGGCGCTGTCG 919  
|||||  
DB 857 CCAAGCTGCCGCCGACCCGGCAGGTACTACGACGCCCGGATGTCACTCGCCTACTAT 916  
QY 920 TGGATCGACGTCCGCCCAACGGCTCG 946  
|||||  
DB 917 TCAATTCGCCGCCGCCCAATCGCGTCG 943

RESULT 12  
AY487552

LOCUS AY487552 1062 bp DNA linear ENV 05-APR-2004  
DEFINITION Uncultured organism clone 2A18 nitriylase (BD5217) gene, complete  
cds.

ACCESSION AY487552  
VERSION AY487552.1 GI:40890308

KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism

REFERENCE unclassified; environmental samples.  
1 (bases 1 to 1062)

AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,B., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.

TITLE Exploring nitriylase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)

REFERENCE 2 (bases 1 to 1062)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,B., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.

AUTHORS Direct Submission  
TITLE Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
JOURNAL Directors Place, San Diego, CA 92121, USA

FEATURES  
source Location/Qualifiers  
1..1062  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A18"  
/environmental\_sample  
1..1062  
/locus\_tag="BD5217"

gene  
1..1062  
/locus\_tag="BD5217"  
/locus\_tag="BD5217"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitriylase"  
/protein\_id="AAR97499.1"  
/db\_xref="GI:40890309"

CDS  
1..1062  
/translation="MSQIHPKLVAAVQAAPAFLLDASIEKTRVYVDEAAAGAKLI  
AFPEWIPGYPMWIMLGAPAWAIMRGFVSRYPDNSLOYGSPPEARLRDAARNKIYIA  
LGLSERDGSLLYIAQWITIGPGSEITVAQRKLPKTHAERTVFGEGDLSHLAVHDLIGR  
LGAICWEHLQPLSKYAMYAONEQIHVAAPSPSLYDPFAHALGAEVNNAAASKIYAVE  
GSCFVIAPCATVSQVMIDELCDTPKHFQFLHVGGFPAVITYGPDGSPPLAKPLPPDQEG  
LYADIDLGMSIVAKAAADPAGHYARPVTRLLFNNRPANREKLALPVDQAEAVDSPL  
KAPDASPRVTKLPSSQAAE"

ORIGIN

Query Match 43.0%; Score 447.8; DB 3; Length 1062;  
Best Local Similarity 67.4%; Pred. No. 3.5e-48;  
Matches 648; Conservative 0; Mismatches 307; Indels 6; Gaps 1;  
QY 28 GCGCGCGCGGTGACAGCGCCCGCGGTGTTCTTCGATCTCGACCGCACAGTCGAGAAAGCG 87  
|||||  
DB 28 GTCGCGGCGGTGACAGCGTGCCTCCGCTTCTCGATCTCGATGTCATCGATCGAAAAAGACA 87  
QY 88 ATCGCGCTGATCGACAGCGCGCCCAAGCAGAGCGTGCCTGATCGCATTCGAGAGACT 147  
|||||  
DB 88 ATACGTATGTGACGAGCGCGGTGCGCGCGCGGCGAAGTTGATTGCGTTCCGGAACCC 147

QY 148 TGGATTCGCCGCTATCCCTTTTGGATATGCGCTGGCGCGCCGGCTTGCGGCAT-----G 201  
|||||  
DB 148 TGGATTCGCCGCTATCCCATGATGATGATGCTCGCGCTCGCGCTCGCGCGCATCATGCGT 207  
QY 202 CGCTTCGTCCAGCCGCTATTTTCGAGAAATTCCTCGTGCGCGGACGACGAGCGCC 261  
|||||  
DB 208 GCGTTGCTTCGCGCTATTTTCGACAACTCTGCTGCAATATGGCAGTCGGAAGCTGAACGG 267  
QY 262 CTGCGGATGCGGCGCCCGCCGACCGCATGTCGTGCGCGGCTATAGCGAGCGCG 321  
|||||  
DB 268 CTGCGGAGCGCGCCGACGCGCAACGATATACATCGCCCTCGGCTGTGCGAGCGGAC 327  
QY 322 GCGGCGAGCTTATATGCGGCGAGCGATCTTGGCCCCGATGCGATCTGATCGCGCG 381  
|||||  
DB 328 GGGGCGAGCTTTATATCGCGCAATGATCATCGGCGCTGCGCGGCAACGCTTGACAA 387  
QY 382 CGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGGTTCGCGCGGAGAGAGCGGACG 441  
DB 388 CGCGCAAGCTCAAGCCACGACGCGCAGCGGCACTGATTCGGCGAAGCGATGTTCA 447  
QY 442 CATCTCGCGGTGACGATACCGGCATCGGCGCGCTCGCGCGCTGTGCTGGAGCAC 501  
DB 448 CATCTGCGCGGTGATGATCTCGATATTTGGAAGATTGGCGCGCTTGTGCTGGGAACAT 507  
QY 502 ATCCAGCCATTTGCAATACGCCATGACGCCCGCGGACGAGCAAGCTTCAAGTCCGCGTCG 561  
DB 508 CTGCAACCGTTGTGCAAAATATGCAATGTACGCCCGCAAGACGAGCAAAATTCACGTCGCGCC 567  
QY 562 TGGCGGAGCTTCAGCCTCTATCGCGGATGCGCTATGCGCTCGGACCGGAGGTCAATACC 621  
DB 568 TGGCGGAGCTTCGCTATACGATCCCTTTGACACGCACTCGCGCGGAGGTCAATACC 627  
QY 622 GCGGCAAGCAGATCTACGCGGTGAGGCGCGCTGCTACGTCGTGCGCGGACCG 681  
DB 628 GCTGCGAGCAGATCTATGCGGTGAGGAGATCGTCTTGTGATGCGCGCGTGCAGACG 687  
QY 682 GTTTCGCGGAGATGATCAAGGTATTTGGTGATACGCCGCAAGAGATGTTCTCTCAAG 741  
DB 688 GTTTCGCGGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 747  
QY 742 GCGGCGGCGGTTTGGCATGATTTTGGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 801  
DB 748 GTCGCGGCGGCTTCGCGCTCATTTACGCTCCGACGCGCTCGCGCAAACTCTT 807  
QY 802 CCGGAGACCGAAGAGGAGCTGCTGCGCGGATATCGACCTCGGATGATCGCGTTGCGC 861  
DB 808 CCGGAGACCGAAGAGGAGCTTCTATGCGGACATGATCTCGGATGATCTCGGCTGCGC 867  
QY 862 AAGCGGCGCGGATCGCGCGGCGCACTATTCACGCGCGGACGTAACGCGCTGCTGCTG 921  
DB 868 AAGCGGCGCGGATCGCGCGGCGCACTATTCACGCTGCGGATGATCTGCGCTGCTGCT 927  
QY 922 GATCGAGTCGCGCGCGCGCGCGCGCTGCTGATGCGCGCATTCGAAACGCAAAACGAG 981  
DB 928 AACATCGCGCGCGCAACCGCGCTGCAAGAGCTGGCGTTGCGCGGTGATCAAGAGCGGAA 987  
QY 982 G 982  
DB 988 G 988

RESULT 13  
AP006618\_34  
WPCOMMENT

Sequence split into 61 fragments LOCUS AP006618 Accession AP006618  
Fragment Name Begin End  
AP006618\_00 1 110000  
AP006618\_01 100001 210000  
AP006618\_02 200001 310000  
AP006618\_03 300001 410000  
AP006618\_04 400001 510000  
AP006618\_05 500001 610000  
AP006618\_06 600001 710000



AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000
AP006618_49	4900001	5010000
AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	6021225

Continuation (35 of 61) of AP006618 from base 3400001 (AP006618 Nocardia farcinica IFM 1

Query Match 42.7%; Score 445; DB 1; Length 110000;  
Best Local Similarity 68.1%; Pred. No. 2.4e-48;  
Matches 619; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

QY	25	CGCGCGCGCGGTGACAGCGCGCGCGGTCTCTCGATCTGACCGCACAGTCGAGAA	84
DB	70070	CGCGCGCGCGGTGACAGCGCGCGCGGTGTGCTGACGCGCGCGACGCTGACAA	70129
QY	85	GCGATCGGCTGATCGAGCAGCGCGCAAGCAGACGTGCGCTGATCGCATTCAGAG	144
DB	70130	TGCGTCGCTTGATCGAGAGCGCGCGACACAGCGCGCGCTGATCGCTTCCCGAG	70189
QY	145	ACTTGATTCCCGGCTATCCCTTTTGATATGGCTGGCGCGCGCGCTTGGGGCATGCG	204
DB	70190	ACGTTCTGCGCGCGCTACCGCTGTGTGTGCTGATTCGCCGCGCGTGGGCATGAC	70249
QY	205	TTGTCGACGCGCTATTTCGAGATTGCTGCTGCGCGCAGCAAGTAGGACGCCCTG	264

DB	70250	TTGTCGCGCGCGCTACTTCGACCACTGCTCGCGCTGACCGCGCGTGTTCGCCCGTTG	70309
QY	265	GCGGATGCGCGCGCGCGCGCGCGCATGCTGCTGCGCGCGCTATAGCAGCGCGCGG	324
DB	70310	CGCGAGCGCGCGCGCGCGCGCATGCTGCTGCGCGCGCGCGCGCGCGCGCGCGG	70369
QY	325	GCGAGCTCTATATGGGCCAGCGCATCTTGGCGCGCGCGCATGCGCATTCGCGCGCG	384
DB	70370	GGTCGCTGTACATGGGCCAGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGG	70429
QY	385	CGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTGTTCGCGCAGGAGAGCGGACCAT	444
DB	70430	CGCAAGCTCAAGCGCGACCCACGTCGAGCGCACCGTCTTCGCTGAGAGCGGCTCGAAC	70489
QY	445	CTCGCGGTGACGATACCGCGCATCGCGCGCGCTCGCGCGCGCTCTGCTGCGGACATC	504
DB	70490	CTGACCGTCTGTGACACCGGAGCTGGCGCGCGCTCGCGCGCGCTGCTGCGGAGCATCTG	70549
QY	505	CAGCCATTGTGAAATACGCCATGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	564
DB	70550	CAGCGCGTGAACCAAGTACCGCATGTACTCCAGCACGAGCAGATCCACGTGGCGGCTGG	70609
QY	565	CCGAGCTTCAGCCTCTATCGCGCGCATGGCCTATGCGCTCGGACCGGAGTCAATACCGG	624
DB	70610	CCGAGCTTCTCGGTCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	70669
QY	625	GCAAGCCAGATCTACGCGGTGACGCGCGCGCGCTGCTACGCTGCGCGCTGCGCGACCGTT	684
DB	70670	GCGGCGCGCGCGCATACGCGGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	70729
QY	685	TCGCGGAGATGATCAAGGTATGTGTGATACGCGCGCGCGCGCGCGCGCGCGCGCGG	744
DB	70730	GACGAGCGCGGTGTGCAATTGTTCTGCGACACACCGCGCGCGCGCGCGCGCGCGCGG	70789
QY	745	GCGCGCGGTTTGGCATGATTTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	804
DB	70790	GCGCGCGGCTTCCGCCAGATCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	70849
QY	805	GAGACCGAAGAGGAGCTGCTGTGTGCGCGCGCATATCGACCTCGGCGATGATCGCGTGGCCAG	864
DB	70850	GAGACCGAAGAGGAGCGCTGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	70909
QY	865	GCGCGCGCGGATCCGCGCGCGCGCATTTACGCGCGCGCGCGCGCGCGCGCGCGCGG	924
DB	70910	AGCGG	70969
QY	925	CGACGTCGG 933	
DB	70970	CCGCGCGCG 70978	

RESULT 14	AY487448	1017 bp	DNA	linear	ENV 05-APR-2004
LOCUS	AY487448				
DEFINITION	Uncultured organism clone 2A25 nitrilase (BD5313) gene, complete cds.				
ACCESSION	AY487448				
VERSION	AY487448.1	GI:40890100			
KEYWORDS	ENV.				
SOURCE	uncultured organism				
ORGANISM	unclassified; environmental samples.				
REFERENCE	1 (bases 1 to 1017)				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chl,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Exploring nitrilase sequence space for enantioselective catalysis				
JOURNAL	Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)				
REFERENCE	2 (bases 1 to 1017)				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,				



RLGALCCWEHLQPLSKYAMYAONEQVHVAAMPFSLYDPFAPALGAENVNNAASRVAV  
EGSCFVLAPCATVSQAMIDELCDRPDKHALLHAGGFAIYGPDSQIEKLAPDQEG  
LLIAEIDLGAIGVAKNAADPAGHXS RPDVTRLLINKKRYQVEQFALPADMVEPADIG  
AAS"

ORIGIN

Query Match 42.1%; Score 438; DB 3; Length 1017;  
Best Local Similarity 68.2%; Pred. No. 6.3e-47;  
Matches 641; Conservative 0; Mismatches 290; Indels 9; Gaps 2;

QY 16 ACGAAGTATCGCGGCGCGGCGGTGACAGCGCGCGGTGTTCTCGATCTCGACCGCACA 75  
Db 16 ACGAAATACAGGTGGCGGTGTGACAGCGCGCGCGCGCTGCTCGACTCGAGGCTCG 75  
QY 76 GTCGAGAAAGCGATCGGCTTGATCGAGCAGCGGCCAAGCAGACGTCGCCCTGATCGCA 135  
Db 76 ATCGGCAAGTCCATCGGCTTAATCAAGAGCGCGCGGACAGAGGCGCCCAAGCTGATCGCC 135  
QY 136 TTCCCAAGACTTGATTCGGGCTATCCCTTTTGATATGCGTGGGCGCGCGCTTGG 195  
Db 136 TTTCCGAGGCTTATCCCGGTTACCCCTGTGTATATCTGATGACTCGCCGCGCTGG 195  
QY 196 -----GGCATGCGCTTCTGTCAGCGCTATTTCGAGAAATTCGCTCGTGGCGCGCAGCAAG 249  
Db 196 GCGATCGGCGCGGCTTCTGTCAGCGCTATTTCGACATTCGCTCTCCTACGACAGTCCC 255  
QY 250 CAGTGGCAGGCGCTTGGCGGATGCGGCGCGCGCGCGCGCGCATGCTATGCGCGGCTAT 309  
Db 256 CAGCGCGAGCGGCTGCGTATGCGGCGCGCGCGCGCGCGCGCATGCTATGCGCGCTG 315  
QY 310 AGCGAGCGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCGCGCGCGCGAT 369  
Db 316 TCCGAACGCGACGCGCGCGCGCGCTTACTGCGCGCAATGTTGATCGGCGCGCGA 375  
QY 370 CTGATCGCGCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGAGCGCGCGTTCGCGGAG 429  
Db 376 ACCATTGCCAAGCGCGCGCGCAAGCTGCGGCGCGCGCGCATGCGGAGCGCGCTATGGCGAA 435  
QY 430 CGAGACGCGCGCGCATCTCGCGGTGCA---CGATACCGCGCATCGGCGCGCTCGCGCGCTC 486  
Db 436 GCGGACGCGCGCGCATCTGCGCGGTACATGCGCGCGCGCGCGCGCATCGGTGCGTGGCGCGCTG 495  
QY 487 TGTGCTGGGAGCACATCCAGCGCATTTGCGAATAACGCCATGTACGCGCGCGCGCGAGACAG 546  
Db 496 TGCTGCTGGGAGCATCTTCAAGCGGTGTGGAAGTACGCAATGTACGCGCGCGCGAGACGAGCAG 555  
QY 547 GTCCACGTGCGGTGCTGCGCGCGCGCGCGCTTATCGCGCGCATGCGCTATGCGCTCGGA 606  
Db 556 GTCCACGTGCGGTGCTGCGCGCGCGCGCGCTTATCGATCCCTTGGCGCGCGCGCTCGGC 615  
QY 607 CCGGAGGTCAATACCGCGCGCAAGCGCGCATCTACCGCGGTGAGGCGCGCTGCTACGTGCTG 666  
Db 616 GCGGAGGTCAACACGCTGCTCGCGCGGTCTATGCGGTGAGGCGCTCTGCTGCTGCTC 675  
QY 667 GCGTGTGCGCGCGACCGTTTTCGCGCGAGATGATCAAGGTATTGAGATACGCCGAGACAAG 726  
Db 676 GCGCTTGGCGCGACGCGGTGTGCGAGCGCATGATCGACGAACCTTGGCATCGCGCGCGATAG 735  
QY 727 GAGATGTTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGCGCGCGCGCGCGCGCC 786  
Db 736 CATGCGCTGCTGATGCGCGCGCGGAGGCTTTTGCAGATCTACGCGCGCGAGCGGAGCCAG 795  
QY 787 CTGGCGGAGCGCGCTCCCGGAGACCGAGAGGAGTGTGCTGCTGCGCGATATCGACCTCGGC 846  
Db 796 ATCGGCGAGAACTGCGCGCGCGGATCAGAGGGTCTGTGATCGCGAGATTGATCTGGGC 855  
QY 847 ATGATCGCGTGGCCCAAGCGCGCGCGCGATCCGCGGCGCGCGCGCGCGCGCGCGCGTA 906  
Db 856 GCCATCGGTGTGCGCAAGACGCGCGAGATCCCGCGGTCAATTAATCACGCGCGCGATGTG 915  
QY 907 ACGCGCTGCTGCTGATCGACGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946  
Db 916 ACGCGGTGCTGCTCAACAGAGCGGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955

Search completed: April 27, 2006, 03:27:23  
Job time : 5622.87 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:18:14 ; Search time 719.835 Seconds  
(without alignments)  
9638.243 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041

Sequence: 1 atgtcggagcccatgacgaa.....gcgccgccgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_21:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*  
14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1041	100.0	1041	4	AAD11166	Aad11166 Nitrilase
2	1041	100.0	1041	10	ADc24116	Adc24116 DNA seque
3	1041	100.0	1041	12	ADH36217	Adh36217 Chemical
4	1041	100.0	1041	12	ADG93916	Adg93916 Nitrilase
5	1041	100.0	1041	12	ADi62515	Adi62515 DNA encod
6	1041	100.0	1041	12	ADi64636	Adi64636 DNA encod
7	682	65.5	1026	10	ADc23766	Adc23766 DNA seque
8	682	65.5	1026	12	ADH35867	Adh35867 Chemical
9	682	65.5	1026	12	ADG93568	Adg93568 Nitrilase
10	682	65.5	1026	12	ADi62165	Adi62165 DNA encod
11	682	65.5	1026	12	ADi64286	Adi64286 DNA encod
12	486	46.7	1062	10	ADc23944	Adc23944 DNA seque
13	486	46.7	1062	12	ADH36045	Adh36045 Chemical
14	486	46.7	1062	12	ADG93746	Adg93746 Nitrilase
15	486	46.7	1062	12	ADi62343	Adi62343 DNA encod
16	486	46.7	1062	12	ADi64464	Adi64464 DNA encod
17	478.2	45.9	1056	10	ADc24074	Adc24074 DNA seque
18	478.2	45.9	1056	12	ADH36175	Adh36175 Chemical
19	478.2	45.9	1056	12	ADG93875	Adg93875 Nitrilase

20	478.2	45.9	1056	12	ADi62473	Adi62473 DNA encod
21	478.2	45.9	1056	12	ADi64594	Adi64594 DNA encod
22	470	45.1	1050	10	ADc24110	Adc24110 DNA seque
23	470	45.1	1050	12	ADH36211	Adh36211 Chemical
24	470	45.1	1050	12	ADG93910	Adg93910 Nitrilase
25	470	45.1	1050	12	ADi62509	Adi62509 DNA encod
26	470	45.1	1050	12	ADi64630	Adi64630 DNA encod
27	457.2	43.9	1065	10	ADc23834	Adc23834 DNA seque
28	457.2	43.9	1065	12	ADH35935	Adh35935 Chemical
29	457.2	43.9	1065	12	ADG93636	Adg93636 Nitrilase
30	457.2	43.9	1065	12	ADi62233	Adi62233 DNA encod
31	457.2	43.9	1065	12	ADi64354	Adi64354 DNA encod
32	455.2	43.7	1077	10	ADc23902	Adc23902 DNA seque
33	455.2	43.7	1077	12	ADH36003	Adh36003 Chemical
34	455.2	43.7	1077	12	ADG93704	Adg93704 Nitrilase
35	455.2	43.7	1077	12	ADi62301	Adi62301 DNA encod
36	455.2	43.7	1077	12	ADi64422	Adi64422 DNA encod
37	454.6	43.7	1017	10	ADc23930	Adc23930 DNA seque
38	454.6	43.7	1017	12	ADH36031	Adh36031 Chemical
39	454.6	43.7	1017	12	ADG93732	Adg93732 Nitrilase
40	454.6	43.7	1017	12	ADi62329	Adi62329 DNA encod
41	454.6	43.7	1017	12	ADi64450	Adi64450 DNA encod
42	453.6	43.6	1017	10	ADc23982	Adc23982 DNA seque
43	453.6	43.6	1017	12	ADH36083	Adh36083 Chemical
44	453.6	43.6	1017	12	ADG93784	Adg93784 Nitrilase
45	453.6	43.6	1017	12	ADi62381	Adi62381 DNA encod

ALIGNMENTS

RESULT 1	
AAD11166	ID AAD11166 standard; DNA; 1041 BP.
XX	XX
AC AAD11166;	XX
XX	XX
DT 25-SEP-2001 (first entry)	XX
XX	XX
DE Nitrilase DNA #1.	DE
XX	XX
Nitrilase; enantiomer; alpha-substituted carboxylic acid;	KW
stereoselective; hydrolysis; amino nitrile; cyanohydrin; ds.	KW
XX	XX
OS Unidentified.	OS
XX	XX
FH Key	FH
FT CDS	FT
FT	FT
FT	FT
XX	XX
PN WO200148175-A2.	PN
XX	XX
PD 05-JUL-2001.	PD
XX	XX
PF 29-DEC-2000; 2000WO-US035555.	PF
XX	XX
PR 29-DEC-1999; 99US-0173609P.	PR
PR 07-DEC-2000; 2000US-0254414P.	PR
XX	XX
PA (DIVE-) DIVERSA CORP.	PA
PA (MADD/) MADDEN D.	PA
XX	XX
PI Madden M, Weiner DP, Chaplin JA;	PI
XX	XX
DR WPI; 2001-465211/50.	DR
DR P-PSDB; AAE05488.	DR
XX	XX
PT Producing alpha-substituted carboxylic acid enantiomers by contacting	PT
aldehyde or ketone with cyanide, ammonia compound or its salt or amine	PT
and hydrolyzing the resulting amino nitrile or cyanohydrin with	PT
Nitrilase.	PT
XX	XX



PS Claim 27; Page 82-84; 87bp; English.  
XX The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolysing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a DNA encoding nitrilase #1  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 1041; DB 4; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGGCGCGCGGTGTTCTC 60  
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGGCGCGCGGTGTTCTC 60  
  
QY 61 GATCTGACCCGACAGTCGAGAAAGCGATGGCCTGATCGAGCAGGCGCCAAAGCAGGAC 120  
DB 61 GATCTGACCCGACAGTCGAGAAAGCGATGGCCTGATCGAGCAGGCGCCAAAGCAGGAC 120  
  
QY 121 GTGCGCCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTGGATATGGCTG 180  
DB 121 GTGCGCCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTGGATATGGCTG 180  
  
QY 181 GGGCGCGCGGCTGGGGCATGCGCTTCGTCACGCGCTATTTCGAATTCGCTCGTGGC 240  
DB 181 GGGCGCGCGGCTGGGGCATGCGCTTCGTCACGCGCTATTTCGAATTCGCTCGTGGC 240  
  
QY 241 GGCAGCAAGCAGTGGCAGAGGCGCGGATGCGGCGCGCGCGCATGATGTCGTG 300  
DB 241 GGCAGCAAGCAGTGGCAGAGGCGCGGATGCGGCGCGCGCGCATGATGTCGTG 300  
  
QY 301 GCCGGCTATAGCAGCGCGCGCGCGCGAGCCTTATATGGCCAGGCGATTCGCGCCC 360  
DB 301 GCCGGCTATAGCAGCGCGCGCGCGCGAGCCTTATATGGCCAGGCGATTCGCGCCC 360  
  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
DB 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
  
QY 421 TTCGGCGAGGAGACGGCAGCCCATCTCCGGGTGCACGATACCCCATCGGGCGCTCGC 480  
DB 421 TTCGGCGAGGAGACGGCAGCCCATCTCCGGGTGCACGATACCCCATCGGGCGCTCGC 480  
  
QY 481 GCGCTCTGTTCTGGGAGCACATCCAGCCATTGTCGAATACGCCATGTACGCCCGCGAC 540  
DB 481 GCGCTCTGTTCTGGGAGCACATCCAGCCATTGTCGAATACGCCATGTACGCCCGCGAC 540  
  
QY 541 GAACAGGTTCCAGTCGCGCTGTCGGCGAGCTTCAGCCTCTATCGGGCATGGCCTATGCG 600  
DB 541 GAACAGGTTCCAGTCGCGCTGTCGGCGAGCTTCAGCCTCTATCGGGCATGGCCTATGCG 600  
  
QY 601 CTCGGACCGGAGGTCAATACCGCCCGAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
DB 601 CTCGGACCGGAGGTCAATACCGCCCGAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
  
QY 661 GTGCTGGCGTGTGCGCGACCGTTTCGCCGAGATGATCAAGGTATTGTGATACGCC 720  
DB 661 GTGCTGGCGTGTGCGCGACCGTTTCGCCGAGATGATCAAGGTATTGTGATACGCC 720  
  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGGCTTTTGCCATGATTTTCGGGCGGACGCG 780  
DB 721 GACAAGAGATGTTCTCAAGCGCGCGGCTTTTGCCATGATTTTCGGGCGGACGCG 780  
  
QY 781 CGCGCCTGGCCGAGCCGCTCCCGAGACCGAAGAGGAGTGTGTCGCCGATATGAC 840  
DB 781 CGCGCCTGGCCGAGCCGCTCCCGAGACCGAAGAGGAGTGTGTCGCCGATATGAC 840  
  
QY 841 CTCGGCATGATCGGTTGGCCAAAGCGCGCGCGATCCGCGCGCACTATTACGGCCC 900

DB 841 CTCGGCATGATCGGTTGGCCAAAGCGCGCGCGATCCGCGCGCACTATTACAGGCC 900  
  
QY 901 GACGTAAACCGGCTGCTGCTGATCGACGTCCGGCCCAACGCGTGTACGCTGATGCC 960  
DB 901 GACGTAAACCGGCTGCTGCTGATCGACGTCCGGCCCAACGCGTGTACGCTGATGCC 960  
  
QY 961 GCATTGGAACCGCAAAACGAGACAGGGCGACGCGCGCGCTGCGGTGTGCGGAA 1020  
DB 961 GCATTGGAACCGCAAAACGAGACAGGGCGACGCGCGCGCTGCGGTGTGCGGAA 1020  
  
QY 1021 AGCGCGCGCGCGCGCAGTAG 1041  
DB 1021 AGCGCGCGCGCGCAGTAG 1041  
  
RESULT 2  
ADC24116  
ID ADC24116 standard; DNA; 1041 BP.  
XX  
AC ADC24116;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA sequence (SeqID 383) encoding a nitrilase enzyme.  
XX gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO200300840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR P-PSDB; ADC24117.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 1; SEQ ID NO 383; 560bp; English.  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 10; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCCGGTGTTCTC 60  
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCCGGTGTTCTC 60  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
DB 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
QY 121 GTGCGCCTGATCGCATTTCCAGAGACTTGATTCGGCTATCCCTTTGGATATGCTG 180  
DB 121 GTGCGCCTGATCGCATTTCCAGAGACTTGATTCGGCTATCCCTTTGGATATGCTG 180  
QY 181 GCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGC 240  
DB 181 GCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGC 240  
QY 241 GGCAGCAGCAGTGGCAGCGCCCTGGCGGATGCGGCCCGCCGACGCGCATGATGCTG 300  
DB 241 GGCAGCAGCAGTGGCAGCGCCCTGGCGGATGCGGCCCGCCGACGCGCATGATGCTG 300  
QY 301 GCGGCTATACGAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
DB 301 GCGGCTATACGAGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
QY 361 GATGCGCATCTGATCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
DB 361 GATGCGCATCTGATCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
QY 421 TTGCGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTCGC 480  
DB 421 TTGCGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTCGC 480  
QY 481 GCGCTCTGTGCTGGAGACATCCAGCATTTGCGAATAACGCCATGTACGCCCGGAC 540  
DB 481 GCGCTCTGTGCTGGAGACATCCAGCATTTGCGAATAACGCCATGTACGCCCGGAC 540  
QY 541 GAACAGATCCACGTCGCGCTGCGCGGCTTACGCTCTATCGCGCATGGCCTATGCG 600  
DB 541 GAACAGATCCACGTCGCGCTGCGCGGCTTACGCTCTATCGCGCATGGCCTATGCG 600  
QY 601 CTCGGACGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
DB 601 CTCGGACGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
QY 661 GTGTCGCGCTGTGCGCGCAGCCGTTTCGCGGAGATGATCAAGGTAATGTGATACGCC 720  
DB 661 GTGTCGCGCTGTGCGCGCAGCCGTTTCGCGGAGATGATCAAGGTAATGTGATACGCC 720  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGGCTTTGCCATGATTTCCGGCCGACGCG 780  
DB 721 GACAAGAGATGTTCTCAAGCGCGCGGCTTTGCCATGATTTCCGGCCGACGCG 780  
QY 781 CGCGCCCTGCGCAGCGCGCTCCCGGAGACCGAAGAGGAGCTGTGTCGCCGATATCGAC 840  
DB 781 CGCGCCCTGCGCAGCGCGCTCCCGGAGACCGAAGAGGAGCTGTGTCGCCGATATCGAC 840  
QY 841 CTGCGCATGATCGCGTTGGCCAAAGCGCGCGGCGATCCGCGGCGCACTATTCAAGCGCC 900  
DB 841 CTGCGCATGATCGCGTTGGCCAAAGCGCGCGGCGATCCGCGGCGCACTATTCAAGCGCC 900  
QY 901 GACGTAACGCGGCTGCTGTGATCGAGCTCCGCGCAACGCGCTGTCAAGCTTGATGCC 960  
DB 901 GACGTAACGCGGCTGCTGTGATCGAGCTCCGCGCAACGCGCTGTCAAGCTTGATGCC 960  
QY 961 GCATTGGAACGCAAAACGAGACAAAGGCGAGCGCGCGGCTGTGTCGGGAA 1020  
DB 961 GCATTGGAACGCAAAACGAGACAAAGGCGAGCGCGCGGCTGTGTCGGGAA 1020  
QY 1021 AGCGCGCGCGCGCGCAGTAG 1041

DB 1021 AGCGCGCGCGCGCGCAGTAG 1041  
RESULT 3  
ADH36217  
ID ADH36217 standard; DNA; 1041 BP.  
XX  
XX ADH36217;  
AC  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase gene sequence SeqID383.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; gene; ds.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI Mcquaid J, Stege J;  
XX  
DR WPI; 2004-142708/14.  
DR P-PSDB; ADH36218.  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 74; SEQ ID NO 383; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a gene which encodes a nitrilase enzyme which can be used in the  
CC method of the invention.  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1041; DB 12; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCCGGTGTTCTC 60  
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCCGGTGTTCTC 60  
QY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
DB 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
QY 121 GTGCGCCTGATCGCATTTCCAGAGACTTGATTCGGCTATCCCTTTGGATATGCTG 180  
DB 121 GTGCGCCTGATCGCATTTCCAGAGACTTGATTCGGCTATCCCTTTGGATATGCTG 180  
QY 181 GCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGC 240

Db 181 GCGCGCCGCTTGCGGCGATGCGCTTCTCCAGCGCTATTTCGAAATTGCTCGTGGC 240  
QY 241 GGCAGCAAGCAGTGGCAGGCGCCCTGGCGGATGCGGCCCGCCAGCGCATGTCTGTG 300  
Db 241 GGCAGCAAGCAGTGGCAGGCGCCCTGGCGGATGCGGCCCGCCAGCGCATGTCTGTG 300  
QY 301 GCCGGCTATAGCAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCGATCTTCGGCCCC 360  
Db 301 GCCGGCTATAGCAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCGATCTTCGGCCCC 360  
QY 361 GATGGCGGATCTGATTCGCCCGCGCGCGCGCAAGCTCAAGCTTACCCTATGCGGAGCGCACCGTG 420  
Db 361 GATGGCGGATCTGATTCGCCCGCGCGCGCGCAAGCTTACCCTATGCGGAGCGCACCGTG 420  
QY 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCAGATACCGCCATCGGCGCCTCGGC 480  
Db 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCAGATACCGCCATCGGCGCCTCGGC 480  
QY 481 GCGCTCTGTGCTGGGAGACACATCCAGCCATTGTGAAATACGCGCATGTACGCCCGCGAC 540  
Db 481 GCGCTCTGTGCTGGGAGACACATCCAGCCATTGTGAAATACGCGCATGTACGCCCGCGAC 540  
QY 541 GAACAGGTCCACGTGCGCTGTGCGCGAGCTTCAAGCTTATCGCGCGATGCGCTATGCG 600  
Db 541 GAACAGGTCCACGTGCGCTGTGCGCGAGCTTCAAGCTTATCGCGCGATGCGCTATGCG 600  
QY 601 CTCGGAACCGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
Db 601 CTCGGAACCGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
QY 661 GTGCTGGCGTGTGCGCGCGCGCGCGCGAGATGATCAAGGTATTGTGATACGCGCC 720  
Db 661 GTGCTGGCGTGTGCGCGCGCGCGCGCGAGATGATCAAGGTATTGTGATACGCGCC 720  
QY 721 GACAAGGAGATGTCTCTCAAGCGCGCGCGGTTTGGCATGATTTCCGGCCGAGCGC 780  
Db 721 GACAAGGAGATGTCTCTCAAGCGCGCGCGGTTTGGCATGATTTCCGGCCGAGCGC 780  
QY 781 CGCGCCCTGGCGCGAGCGCGCTCCCGGAGACCGAAGAGGAGACTGTGCTGCGCGATATCGAC 840  
Db 781 CGCGCCCTGGCGCGAGCGCGCTCCCGGAGACCGAAGAGGAGACTGTGCTGCGCGATATCGAC 840  
QY 841 CTCGGCATGATCGCGTTGGCAAGCGCGCGCGCGATCCGCGCGGCACTATTACCGGCC 900  
Db 841 CTCGGCATGATCGCGTTGGCAAGCGCGCGCGCGATCCGCGCGGCACTATTACCGGCC 900  
QY 901 GACGTAAAGCGGCTGTCTGTGATCGACGTCCGCGCCCAACGCGTTCGTACAGCTTGATGCC 960  
Db 901 GACGTAAAGCGGCTGTCTGTGATCGACGTCCGCGCCCAACGCGTTCGTACAGCTTGATGCC 960  
QY 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGCGCGCTGCGCGTGTGGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGCGCGCTGCGCGTGTGGCGGAA 1020  
QY 1021 AGCGCGCGCGCGCGCGAGTAG 1041  
Db 1021 AGCGCGCGCGCGCGCGAGTAG 1041

RESULT 4  
ADG93916  
ID ADG93916 standard; DNA; 1041 BP.

XX ADG93916;  
AC  
XX  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitriylase enzyme gene sequence SeqID383.  
XX  
KW nitriylase; nitriyl; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; gene; ds.  
XX  
OS Unidentified.

XX  
PN WO2003097810-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-090637/09.  
DR P-PSDB; ADG93917.  
XX  
PT New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitriylase activity, useful for screening enantioselective transformation.  
XX  
PS Claim 44; SEQ ID NO 383; 295pp; English.  
XX

CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitriylase activity. Nitriylase's are capable of  
CC converting nitriyl's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitriylase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitriylase enzyme of the invention.  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 12; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTCCGAGGCCCATGACGAAATATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60  
Db 1 ATGTCCGAGGCCCATGACGAAATATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60  
QY 61 GATCTCGAACCGCACAGTGCAGAAAGCGATCGGCTGTATCGACAGCGCGCCAGCAGGAC 120  
Db 61 GATCTCGAACCGCACAGTGCAGAAAGCGATCGGCTGTATCGACAGCGCGCCAGCAGGAC 120  
QY 121 GTGCGCTGATCGCATTCCTCCAGAGACTTGATTCGCGGCTATCCCTTTGATATGCTG 180  
Db 121 GTGCGCTGATCGCATTCCTCCAGAGACTTGATTCGCGGCTATCCCTTTGATATGCTG 180  
QY 181 GCGCGCGCGGCTTGGGGATGCGCTTGTCTCAGCGCTATTTCGAGAAATTCGCTGCGC 240  
Db 181 GCGCGCGCGGCTTGGGGATGCGCTTGTCTCAGCGCTATTTCGAGAAATTCGCTGCGC 240  
QY 241 GGCAGCAAGCAGTGGCAGGCGCCCTGGCGGATGCGCGCGCGCCAGCGCATGTCGTG 300  
Db 241 GGCAGCAAGCAGTGGCAGGCGCCCTGGCGGATGCGCGCGCGCCAGCGCATGTCGTG 300  
QY 301 GCCGGCTATAGCAGCGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCGATCTTCGGCCCC 360  
Db 301 GCCGGCTATAGCAGCGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCGATCTTCGGCCCC 360  
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCCTATGCGGAGCGCACCGTG 420  
Db 361 GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTTACCCTATGCGGAGCGCACCGTG 420  
QY 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCAGATACCGCCATCGGCGCCTCGGC 480  
Db 421 TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCAGATACCGCCATCGGCGCCTCGGC 480  
QY 481 GCGCTCTGTGCTGGGAGACACATCCAGCCATTGTGAAATACGCGCATGTACGCCCGCGAC 540  
Db 481 GCGCTCTGTGCTGGGAGACACATCCAGCCATTGTGAAATACGCGCATGTACGCCCGCGAC 540



QY 541 GAACAGGTCCAAGTCCGCTGCTGCGCCGAGCTTACGCTCTATCCGCGCATGGCCTATGCG 600  
| | | | |  
Db 541 GAACAGGTCCAAGTCCGCTGCTGCGCCGAGCTTACGCTCTATCCGCGCATGGCCTATGCG 600  
QY 601 CTGGAGCCGGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTAC 660  
| | | | |  
Db 601 CTGGAGCCGGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTAC 660  
QY 661 GTGCTGGCGTCTGCGCGAGCCGCTTTCGCCGAGATGATCAAGGTATTGTGATACGCCC 720  
| | | | |  
Db 661 GTGCTGGCGTCTGCGCGAGCCGCTTTCGCCGAGATGATCAAGGTATTGTGATACGCCC 720  
QY 721 GACAAGAGATGTTCTCTCAAGCGCGCGGCTTTTGCCATGATTTCCGGCCCGACGCG 780  
| | | | |  
Db 721 GACAAGAGATGTTCTCTCAAGCGCGCGGCTTTTGCCATGATTTCCGGCCCGACGCG 780  
QY 781 CGCGCCCTGGCCGAGCCGCTCCCGAGACCCAAAGGAGCTGTGTGCGCCGATATCGAC 840  
| | | | |  
Db 781 CGCGCCCTGGCCGAGCCGCTCCCGAGACCCAAAGGAGCTGTGTGCGCCGATATCGAC 840  
QY 841 CTGGGATGATCGCGTGTGGCCAAAGCGCGCGGATCCGGCGGGCCACTATTCAAGCGCC 900  
| | | | |  
Db 841 CTGGGATGATCGCGTGTGGCCAAAGCGCGCGGATCCGGCGGGCCACTATTCAAGCGCC 900  
QY 901 GACGTAACGCGCTGCTGCTGATCGACGTCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
| | | | |  
Db 901 GACGTAACGCGCTGCTGCTGATCGACGTCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
QY 961 GCATTCGAACCGCAAAACGAGGACCAAGGGCGACGCGCCGCTGCGGTGGTGGCGAA 1020  
| | | | |  
Db 961 GCATTCGAACCGCAAAACGAGGACCAAGGGCGACGCGCCGCTGCGGTGGTGGCGAA 1020  
QY 1021 AGCGCCCGCCGCGCGCAGTAG 1041  
| | | | |  
Db 1021 AGCGCCCGCCGCGCGCAGTAG 1041  
RESULT 5  
ADI62515  
ID ADI62515 standard; DNA; 1041 BP.  
XX  
AC ADI62515;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE DNA encoding nitrilase polypeptide #192.  
XX  
KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipemic; gene; ds.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
DR WPI; 2004-090821/09.  
DR P-PSDB; ADI62516.  
XX

PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 46; SEQ ID NO 383; 253pp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence encodes a nitrilase polypeptide obtained from an environmental  
CC sample.  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 12; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCGCGGTGTTCTC 60  
| | | | |  
Db 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGTGACAGCCGCGGTGTTCTC 60  
QY 61 GATCTGCACCGCACAGTCGAGAAACGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
| | | | |  
Db 61 GATCTGCACCGCACAGTCGAGAAACGATCGGCTGATCGAGCAGCGGCCAAGCAGAC 120  
QY 121 GTGGCGCTGATCGCATTCGCCAGACTTGATTCGCCGCTATCCCTTTTGATATGCGTG 180  
| | | | |  
Db 121 GTGGCGCTGATCGCATTCGCCAGACTTGATTCGCCGCTATCCCTTTTGATATGCGTG 180  
QY 181 GCGCGCGGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAGAATTGCTCGTGCGC 240  
| | | | |  
Db 181 GCGCGCGGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAGAATTGCTCGTGCGC 240  
QY 241 GGCAGCAGCAGTGGCAGGCGCTGCGGATGCGGCCCGCCGCAAGCATGATGTCGTG 300  
| | | | |  
Db 241 GGCAGCAGCAGTGGCAGGCGCTGCGGATGCGGCCCGCCGCAAGCATGATGTCGTG 300  
QY 301 GCCGCTATTAGCGAGCGCGCGGCGGAGCCTCTATATGGGCCAGCGCATCTTCGGCCCC 360  
| | | | |  
Db 301 GCCGCTATTAGCGAGCGCGCGGCGGAGCCTCTATATGGGCCAGCGCATCTTCGGCCCC 360  
QY 361 GATGGCATCTGATCGCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGTG 420  
| | | | |  
Db 361 GATGGCATCTGATCGCGCGCGCGCAAGCTCAAGCCTTACCATGCGGAGCGCACCGTG 420  
QY 421 TTGGCGAGGAGACGCGCAGCCATCTCGCGTGACAGATACCGCATCGGGCGCTCGGC 480  
| | | | |  
Db 421 TTGGCGAGGAGACGCGCAGCCATCTCGCGTGACAGATACCGCATCGGGCGCTCGGC 480  
QY 481 GCGCTCTGTTGCTGGAGCAATCCAGCATTTGTGAATAACGCATGTACGCGCGGAC 540  
| | | | |  
Db 481 GCGCTCTGTTGCTGGAGCAATCCAGCATTTGTGAATAACGCATGTACGCGCGGAC 540  
QY 541 GAACAGGTCCAAGTCCGCTGTCGCCGAGCTTACGCTCTATCGCGCATGCGCTATGCG 600  
| | | | |  
Db 541 GAACAGGTCCAAGTCCGCTGTCGCCGAGCTTACGCTCTATCGCGCATGCGCTATGCG 600  
QY 601 CTGGAGCCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGCGGTGAGGGCGGCTGCTAC 660  
| | | | |  
Db 601 CTGGAGCCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGCGGTGAGGGCGGCTGCTAC 660



QY	661	GTGCTGGCGTCTGTGCGCGACCGGTTTCGCCGAGATGATCAAGGTATTGGTGATACGCCC	720
Db	661	GTGCTGGCGTCTGTGCGCGACCGGTTTCGCCGAGATGATCAAGGTATTGGTGATACGCCC	720
QY	721	GACAAGAGATGTTTCTCTCAAGGCGGCGGGTTTGGCCATGATTTTGGGGCCCGACGGC	780
Db	721	GACAAGAGATGTTTCTCTCAAGGCGGCGGGTTTGGCCATGATTTTGGGGCCCGACGGC	780
QY	781	CGGCGCCTGGCGGAGCCGCTCCCGAGACCGAAGGAGTGTGTGCGCCGATATCGAC	840
Db	781	CGGCGCCTGGCGGAGCCGCTCCCGAGACCGAAGGAGTGTGTGCGCCGATATCGAC	840
QY	841	CTGGGATGATCGCGTTGGCCAAAGCGCGCGGATCCGGCGGGCCATATTCACGGCCC	900
Db	841	CTGGGATGATCGCGTTGGCCAAAGCGCGCGGATCCGGCGGGCCATATTCACGGCCC	900
QY	901	GACGTAACGGCGCTGCTGCTGGATGACGTCGCGGCCCAACGCGTCGTCACGCTTGATGCC	960
Db	901	GACGTAACGGCGCTGCTGCTGGATGACGTCGCGGCCCAACGCGTCGTCACGCTTGATGCC	960
QY	961	GCATTGGAACCGCAAAACGAGACAAGGGCGACGCGCCGCGTGGCGGTGGCGGAA	1020
Db	961	GCATTGGAACCGCAAAACGAGACAAGGGCGACGCGCCGCGTGGCGGTGGCGGAA	1020
QY	1021	AGCGCGCGCGCGCGCAGTAG 1041	
Db	1021	AGCGCGCGCGCGCGCAGTAG 1041	
RESULT 6			
AD164636 standard; DNA; 1041 BP.			
ID	AD164636		
XX	AD164636;		
AC	AD164636;		
XX	22-APR-2004 (first entry)		
DT	22-APR-2004 (first entry)		
XX	DNA encoding nitrilase seq id 192.		
DE	(R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;		
XX	carboxylic acid; cyanohydrin moiety hydrolysis;		
KW	aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;		
KW	chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;		
KW	(R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;		
KW	(R)-phenyl lactic acid derivative; & enantiomeric excess;		
KW	& diastereomeric excess; food additive; drug intermediate; ds; nitrilase;		
KW	gene.		
OS	Unidentified.		
XX	US2004014195-A1.		
PN	22-JAN-2004.		
XX	15-MAY-2003; 2003US-00440523.		
PF	29-DEC-1999; 99US-0173609P.		
XX	07-DEC-2000; 2000US-0254414P.		
PR	28-DEC-2000; 2000US-00751299.		
PR	21-JUN-2001; 2001US-0300189P.		
PR	30-JUL-2001; 2001US-0309006P.		
PR	22-JAN-2002; 2002US-0351336P.		
PR	15-MAY-2002; 2002US-00146772.		
PR	09-SEP-2002; 2002US-00241742.		
XX	(DIVE-) DIVERSA CORP.		
PA	Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;		
XX	WPI; 2004-121569/12.		
DR	P-PSDB; AD164637.		
DR	Novel isolated or recombinant polypeptide having nitrilase activity,		
XX			
PT			

PT	useful in production of food additives.
XX	
PS	Claim 1; SEQ ID NO 383; 105pp; English.
XX	
CC	The invention describes an isolated or recombinant polypeptide (I)
CC	comprising amino acids having a sequence at least 50 % identical to a
CC	sequence (S1) available in electronic form (EC) from the following web
CC	site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its
CC	variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC	residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC	fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC	hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC	hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
CC	chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC	producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC	mandelic acid or an (R)-phenyl lactic acid derivative; producing an (R)-
CC	derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
CC	and for identifying a modified compound. The inventive method is useful
CC	for monitoring or determining % enantiomeric excess or % diastereomeric
CC	excess. (I) is useful in the production of food additives and drug
CC	intermediates. This sequence encodes a nitrilase of the invention.
XX	
XX	
Seq	Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;
QY	Query Match 100.0%; Score 1041; DB 12; Length 1041;
Best	Best Local Similarity 100.0%; Pred. No. 1.1e-197;
Matches	Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db	1 ATGTCGAGACCCATGACGAAATATCGCGCGCGGCTGCAAGCGCGGTTTCCTC 60
QY	1 ATGTCGAGAGCCCATGACGAAATATCGCGCGCGGCTGCAAGCGCGGTTTCCTC 60
Db	61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCGCAAGCAGAC 120
QY	61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCGCAAGCAGAC 120
Db	61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGGCGCAAGCAGAC 120
QY	121 GTGGGCTGATCGCATTCGACAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTG 180
Db	121 GTGGGCTGATCGCATTCGACAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTG 180
QY	181 GCGCGCGCGGCTTGGGGATGCGCTTGTCCAGCGCTATTTCGAGAAATTCGTCGTGCGC 240
Db	181 GCGCGCGCGGCTTGGGGATGCGCTTGTCCAGCGCTATTTCGAGAAATTCGTCGTGCGC 240
QY	241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGATGCGGCGCGCCGCAAGCATGATGTGCTG 300
Db	241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGATGCGGCGCGCCGCAAGCATGATGTGCTG 300
QY	301 GCCGGCTATAGCAGCGCGCGCGGCGAGCGCTTATATGGCCAGCGCATTCGGCCCC 360
Db	301 GCCGGCTATAGCAGCGCGCGCGGCGAGCGCTTATATGGCCAGCGCATTCGGCCCC 360
QY	361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGACCGTG 420
Db	361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGACCGTG 420
QY	421 TTCGGCAGGAGAGACCGGAGCCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGC 480
Db	421 TTCGGCAGGAGAGACCGGAGCCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGC 480
QY	481 GCGCTCTGTGCTGGAGCACATCCAGCCATTGTCGAATAACGCCATGTACGCGCGCCGAC 540
Db	481 GCGCTCTGTGCTGGAGCACATCCAGCCATTGTCGAATAACGCCATGTACGCGCGCCGAC 540
QY	541 GAACAGGTCCACGTCGCGTCTGTGGCCGAGCTTCAAGCCTTATCGCGGCATGGCTATGCG 600
Db	541 GAACAGGTCCACGTCGCGTCTGTGGCCGAGCTTCAAGCCTTATCGCGGCATGGCTATGCG 600
QY	601 CTCGGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
Db	601 CTCGGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
QY	661 GTGCTGGCGTCTGTGCGCGACCGTTTCGCCGAGATGATCAAGGTATTGGTGAATACGCC 720

|||||  
DB 661 GTGTCGGCTGTCGCCGACCGTTTCGCCGAGATGATCAAGGTATTGTGATACGCC 720  
QY 721 GACAAGAGATGTTCTCAAGGCCGGCGGTTTGGCATGATTTCCGGCCGACGGC 780  
DB 721 GACAAGAGATGTTCTCAAGGCCGGCGGTTTGGCATGATTTCCGGCCGACGGC 780  
QY 781 CGCGCCCTGGCCGAGCCGCTCCCGAGACCGAAGGAGTGTGTGCGCGATATCGAC 840  
DB 781 CGCGCCCTGGCCGAGCCGCTCCCGAGACCGAAGGAGTGTGTGCGCGATATCGAC 840  
QY 841 CTCGGCATGATCGGTTGGCCAAGCGCGCGCGATCCGGCGGCCACTATTCACGGCCC 900  
DB 841 CTCGGCATGATCGGTTGGCCAAGCGCGCGCGATCCGGCGGCCACTATTCACGGCCC 900  
QY 901 GACGTACCGCGCTGCTGTGGATCGACGTCCGGCCCAACGGCTGTACGCTTGATGCC 960  
DB 901 GACGTACCGCGCTGCTGTGGATCGACGTCCGGCCCAACGGCTGTACGCTTGATGCC 960  
QY 961 GCATTGCAACCGCAAAAGAGACAAGGGCGACGGCCCGCTGCGCGTGTGGCGGAA 1020  
DB 961 GCATTGCAACCGCAAAAGAGACAAGGGCGACGGCCCGCTGCGCGTGTGGCGGAA 1020  
QY 1021 AGCGCCGCGCGCGCGAGTAG 1041  
DB 1021 AGCGCCGCGCGCGCGAGTAG 1041

RESULT 7  
ADC23766  
ID ADC23766 standard; DNA; 1026 BP.  
XX AC ADC23766;  
XX DT 18-DEC-2003 (first entry)  
XX DE DNA sequence (SeqID 33) encoding a nitrilase enzyme.  
XX KM gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX OS enantiomer; chiral medicine.  
XX OS Unidentified.  
PN WO2003000840-A2.  
XX PD 03-JAN-2003.  
XX PF 15-MAY-2002; 2002WO-US015983.  
XX PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX DR WPI; 2003-201417/19.  
DR P-PSDB; ADC23767.  
XX PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX PS Claim 1; SEQ ID NO 33; 560bp; English.  
XX CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and

CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 10; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTGAGGCCCATGACGAAGTATCGCGCGCGGTGACAGGCCGCGGTTCCTC 60  
DB 1 ATGTTAAGTCCGTGACGAGTATCGCGCGCGGTGACAGGCCGCGGTTCCTC 60  
QY 61 GATCTGACCGCACAGTGCAGAAAGCGATCGCCTGATCGAGCAGCGCCAGCAGAC 120  
DB 61 GATCTGACCGCACCGTGCAGAAAGCGATCGCATCATGACAGCGCGCCAGCAGAT 120  
QY 121 GTGCGCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCTG 180  
DB 121 GTGCGCTGATCGCGTTTCCGGAACCTTGATTCGCCGCTATCCGCTTGATCTGCTC 180  
QY 181 GCGCGCCGCGCTTGGGGCATGCGCTTGTCCAGCGCTATTTCGAATTCGCTGTCGC 240  
DB 181 GCGTGGCGGCTGGGGCATGCGCTTGTCCAGCGCTATTTCGAATTCGCTGTCGC 240  
QY 241 GGCAGCAAGCAGTGGCAGGCCCTGCGGATCGGCCCGCGCCAGCGATGATGCTG 300  
DB 241 GGCAGCAACAGTGAACCGGATCGCGGATCGGCCCGCGCCAGCGATGACCGTCTC 300  
QY 301 GCCGCTATAGCGAGCGCGCGCGGCACTCTATATGAGCGAGGATCTTCGCCCC 360  
DB 301 GTCGCTTACGAGCGCGCGGAGCGGCTCTACATGAGCGAGGATCTTCGCCCC 360  
QY 361 GATGCGATCTGATCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
DB 361 GAAGGCGAGCTCATCGCGCGCGCGGAGAGCTCAAGCCGACACAGCGGAGCGAGCGTG 420  
QY 421 TTCGGGAGGAGACGGCAGCCATCTGCGGTGACAGTACCGCCATCGGCGCTCGGC 480  
DB 421 TTCGGGAGGAGACGGCAGCCATCTGCGGTGACAGAGCGGCGTGTGCTGCTCGGC 480  
QY 481 GCGCTCTGTTCTGGGAGACATCCAGCATTTGCAATACGCCATGTACGCCCGGAC 540  
DB 481 GCCCTCTGCTGGGAGACATCCAGCGCTCTGAAATACGCGATGTATCGGCCAAC 540  
QY 541 GAACAGGTCCAGTCCGCTGTCGTCGAGCCTTCAAGCCTTATCGCGCATGAGCCTATGCG 600  
DB 541 GAACAGGTGATGTGCGCTGTCGTCGAGCCTTCAAGCCTTATCGCGCATGAGCCTATGCG 600  
QY 601 CTCGACCGGAGTCAATACCGCGGACAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660  
DB 601 CTCGAGCGGAGGTGAACACCGCGGAGCCAGGTCTACGCGGTGAGGGCGGCTGTAC 660  
QY 661 GTGCTGGCGTCTGCGCGACCGTTTCCCGGAGATGATCAAGGTATTGTGATACGCC 720  
DB 661 GTGCTGGCGTCTGCTGTCGTGACACCTGAGATCTGAAGGTGCTGATCGACACGCC 720  
QY 721 GACAAGAGATGTTCTCAAGGCCGGCGGCTTTGGCATGATTTGGGCCGACGGC 780  
DB 721 GACAAGAGCGCTGCTGCTGCGCGGGGGTCTGATGATCTTCGCCCGGACGGC 780  
QY 781 CGCGCCCTGGCCGAGCGGCTCCCGAGACGAAGAGGAGTGTGTGCGCGATATCGAC 840  
DB 781 CGCGCGCTGCGCCAGCGGCTGCGGAGACGAAGAGGAGTGTGTGCGCGAGATCGAT 840  
QY 841 CTCGCATGATCGCGTTGGCCAAGCGCGCGCGATCCGCGGCGCACTATTACAGGCC 900





KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; gene; ds.  
XX Unidentified.  
OS  
XX WO2003097810-A2.  
XX PD 27-NOV-2003.  
XX PF 15-MAY-2003; 2003WO-US015712.  
XX PR 15-MAY-2002; 2002US-00146772.  
XX PR 09-SEP-2002; 2002US-00241742.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX WPI; 2004-090637/09.  
DR P-PSDB; ADG93569.  
XX  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
PS Claim 44; SEQ ID NO 33; 295bp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 12; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

OY 1 ATGTCGAGACCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTCCTC 60  
DB 1 ATGTTAAGTCCCGTAGCAGATATCGCGCGCGCGGTGACAGCGCGCGCATCTTTCTC 60  
OY 61 GATCTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAC 120  
DB 61 GATCTCGACCGCACCGTCGAGAAAGCAGATCGCATCATGAGCAGCGCGCAGCAGAGAT 120  
OY 121 GTGCGCCTGATCGCATTCACAGAGACTTGATTCGCCGCTATCCCTTTGATATGCGCTG 180  
DB 121 GTGCGCCTGATCGCGCTTCCGGAACCTGATTCGCCGCTATCCGCTGATCTGCGCTC 180  
OY 181 GCGCGCGCGGCTTGGGCGATGCGCTTCGTCCAGCGCTATTTCGAGAATTCCGTCGCGC 240  
DB 181 GGCTCGCCGGCCTGGGCGATGCGCTTCGTGACGCGCTATTTCGAGAACTCGCTGTGCGC 240  
OY 241 GGCAGCAAGCAGTGGCAGGCGCTTGCGGATGCGGCGCGCGCCAGCAGCATGATGTCGTG 300  
DB 241 GGCAGCAAAACAGTGAACGCGATCGCCGATGCGGCGCGCCACCGCATGACCGTCTGTC 300  
OY 301 GCCGCTATAGCAGCGCGCGCGCGCGCAGCCTCTATATGAGCCAGGCGATCTTGGCCCC 360  
DB 301 GTCGGCTTACGCGAGCGCGGGAGGAGCCTCTACATGAGGCCAGGCGATCTTGGCCCC 360  
OY 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTAACCCATGCGAGCGCACCGTG 420  
DB 361 GAAGCGAGCTCATCGCGCGCGCGCGGAAGCTCAAGCCGACACAGCGCGAAGCGGTG 420  
OY 421 TTCGGCGAGGAGCAGCGCAGCCATCTCGCGGTGCAACGATACCGCCATCGGCGCGCTCGGC 480  
DB 421 TTCGGCGAGGCGCAGCGCAGCCACTTGGCCGTTTACGAGACGGGCGTGTGTCGATCGGC 480

OY 481 GCGCTGTGTTGCTGGAGACACATCCAGCCATTGTCGAAATACGCCCATGACCGCCGAC 540  
DB 481 GCCCTGTGCTGGAGACACATCCAGCCGCTCTGAAATACCGCATGATGCGGCCAAC 540  
OY 541 GAACAGGTCCACGTCGCGTCGTGAGCCGAGCTTACGCTTATCGCGCATGCGCTATGCG 600  
DB 541 GAACAGGTGATGTGCGCTCGTGCCTGCGCGCTTACGCTTATCGCGCATGCGCTATGCG 600  
OY 601 CTCGACCGGAGGTCAATACCGCGCAAGCCAGATCTACGCGGTGAGCGCGGCTGCTAC 660  
DB 601 CTCGGCGCGGAGGTGAACACCGCGCGAGCAGCTACGCGGTGAGGGCGGCTGCTAC 660  
OY 661 GTGCTGGCGGTGCGCGCGCACCGCTTCCGCCGAGATGATCAAGGTATTGTTGATACGCC 720  
DB 661 GTGCTGGCTCTCTGTCTGTCGTGACACCCGAGATCCTGAAGGTGCTGATCGACACGCC 720  
OY 721 GACCAAGAGATGTTCTCAAGGCGCGCGGCTTTGCCATGATTTGGGCGCCGACGGC 780  
DB 721 GACCAAGAGCGCTTGTCTGCTCGCGCGCGGGGGTTCTGATGATCTTGGCGCCGACGGC 780  
OY 781 CGCGCCCTGCGGACCGCGCTCCCGGAGACCGAAGAGGACTGTGTCGCGCATATCGAC 840  
DB 781 CGCGCGCTGCGCCAGCGCGCTGCGGAGACCGAAGAGGGCTCGTACGCGCGAGATCGAT 840  
OY 841 CTCGCGATGATCGGCTTGGCCAAGGCGCGCGCGATCCGCGGCGCACTATTCACGGCCC 900  
DB 841 CTCGGCGGATCGCGCTCGCCAGAGCGCGCGCGCATCCGCGCGCATTCGCGCGGCC 900  
OY 901 GACGTACCGCGCTGCTGCTGATCGACGTCGCGGCCCAACGCGCTGTCACGCTTGATGCC 960  
DB 901 GACGTACCGCGGTTGTTGCTGTAACCCGCGCGCGCGCGCGCTGTAAGCGCTGGTCCG 960  
OY 961 GCATTGGAACCGCAAAACGAGACAAAGGCGACGCGCGCGCTGCGCGGTGCGCGG 1018  
DB 961 CGCTTGAGGTGCTGACAGAGCGAGCAGCGCGCGCGCCACGCAACCGCGCAAGCGG 1018

RESULT 10  
AD162165  
ID AD162165 standard; DNA; 1026 BP.  
XX

AC AD162165;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE DNA encoding nitrilase polypeptide #17.  
XX  
XX Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipaeimic; gene; ds.  
XX  
OS Unidentified.  
XX  
XX WO2003106415-A2.  
XX PD 24-DEC-2003.  
XX  
XX PF 13-JUN-2003; 2003WO-US018840.  
XX  
XX PR 13-JUN-2002; 2002US-0389317P.  
XX PR 28-JUN-2002; 2002US-0392944P.  
XX  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
XX WPI; 2004-090821/09.  
DR P-PSDB; AD162166.  
XX



PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.

PS Claim 46; SEQ ID NO 33; 253bp; English.

XX The present invention relates to a method for preparing an atorvastatin  
XX intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate processes, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence encodes a nitrilase polypeptide obtained from an environmental  
CC sample.

XX Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 12; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

OY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGGTGCAGGCCGCGGTGTCTCTC 60  
DB 1 ATGTTAAGTCCCGTAGCGAGTATCGCGCGCGGGTGCAGGCCGCGCATCTTTCTC 60  
OY 61 GATCTGACCGCAGATCGAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGAGAC 120  
DB 61 GATCTGACCGCAGCCTGAGAAAGCGATCGAGATCATCGAGCAGCGGCCGAGCAGAT 120  
OY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATCCCGGCTATCCCTTTGGATATGCTG 180  
DB 121 GTGCGCTGATCGCTTTCCGAAACCTGATTTCCGGCTATCCGCTTGATCTGCTC 180  
OY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTCGAGAATTCGCTCGTGGC 240  
DB 181 GCGTCCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTCGAGAATTCGCTCGTGGC 240  
OY 241 GGCAGCAAGCAGTGGCAGCGCCCTGCGGATGCGGCCCGCCGACGCATGATGCTGTG 300  
DB 241 GGCAGCAACAGTGGAAACGCGATCGCGCATGCGCGCGCCGACGCATGACCGTCTC 300  
OY 301 GCCGCTATAGCAGCGCGCGCGCGCCTCTATATGGGCCAGCGCATCTTCGCCCC 360  
DB 301 GTCGCTTCAGCAGCGCGCGCGGAGCGCCTCTACATGGGCCAGCGCATCTTCGCCCC 360  
OY 361 GATGCGATCTGATCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420  
DB 361 GAAAGCGAGCTCATCGCGCGCGCGGAACTCAAGCCGACACACGCCGAGCGAAGCGTG 420  
OY 421 TTGCGCAGGAGAGACGCGCAGCCATTCGCGGTGCAGATACCGCCATCGGGCGCTCGG 480  
DB 421 TTGCGCAGGAGGAGACGCGCAGCCATTCGCGGTTCAGAGACGGGCGTGTGTCGATCGG 480  
OY 481 GCGCTCTGTGCTGGAGACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCAC 540  
DB 481 GCCCTCTGCTGCTGGAGACATCCAGCCGCTCTCGAATACCGCATGTATGCCGCCAAC 540  
OY 541 GAACAGGTCCACGTGCGTGTGCGGAGCTTCAGCCTCTATCGCGGCATGCGCTATGCG 600  
DB 541 GAACAGGTGCATGTGCGCTGTGCGGCTTCAGCCTTTATCGCGGCATGCGCTATGCG 600  
OY 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACCGCGTGAAGGCGGCTGCTAC 660  
DB 601 CTCGGCCCGGAGGTGAACACCGCCCGCAGCCAGGTCTACCGCGTGAAGGCGGCTGCTAC 660

OY 661 GTGCTGCGGTCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGTATTGGTGATACGCC 720  
DB 661 GTGCTGCGGTCTGTCGTCGTCGTGACACCCGAGATCTGAAGGTGCTGATGACACGCC 720  
OY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTTGGCATGATTTTGGGCCCGACGCG 780  
DB 721 GACAAGAGCGCTTGTGCTGCTCGCGCGCGGGGCTTCTCGATGATCTTCGCCCGACGCG 780  
OY 781 CGCGCCCTGAGCCGAGCCGCTCCCGGAGACCGAAGAGGACTGCTGTCGCCGATATGAC 840  
DB 781 CGCGCGCTCGCCAGCCGCTGCGGAGACCGAAGAGGAGCTGTCACGCGCGAGATCGAT 840  
OY 841 CTCGCGCATGATCGCGTGTGGCCAGGCGCGCGCGCATCCGCGCGGCCACTATTACGCGCC 900  
DB 841 CTCGCGCGATCGCGCTGCTGCGCCAGGCGCGCGCGCATCCGCGCGCCATTACGCGCGCC 900  
OY 901 GACGTACGCGGCTGCTGATCGACGTCGCGGCCCAACGCGCTGTCACGCTTGATGCC 960  
DB 901 GACGTACGCGGCTTGTGCTGAACCGCGCGGCCCGCGCGCGCTGGAAGCGCTGGTCCG 960  
OY 961 GCATTGCAACCGCAAAACGAGCAAGGGCGACGCGCGCCGCTGCGGCTGTGGCGG 1018  
DB 961 CGCTTGAGGTCTGTGACAGAGCGAGCGCGGAGCGCCCGCCAGCAACCGCGCGAAGCGG 1018

RESULT 11  
AD164286  
ID AD164286 standard, DNA; 1026 BP.

XX AC AD164286;  
XX XX  
DT 22-APR-2004 (first entry)

XX DNA encoding nitrilase seq id 17.

XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KW carboxylic acid; cyanohydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KW (R)-phenyl lactic acid derivative; % enantiomeric excess;  
KW % diastereomeric excess; food additive; drug intermediate; ds; nitrilase;  
KW gene.

XX Unidentified.

XX OS  
XX US2004014195-A1.

XX PD  
XX 22-JAN-2004.

XX PF  
XX 15-MAY-2003; 2003US-00440523.

XX PR  
XX 29-DEC-1999; 99US-0173609P.

XX PR  
XX 07-DEC-2000; 2000US-0254414P.

XX PR  
XX 28-DEC-2000; 2000US-00751299.

XX PR  
XX 21-JUN-2001; 2001US-0300189P.

XX PR  
XX 30-JUL-2001; 2001US-0309066P.

XX PR  
XX 22-JAN-2002; 2002US-0351336P.

XX PR  
XX 15-MAY-2002; 2002US-00146772.

XX PR  
XX 09-SEP-2002; 2002US-00241742.

XX PA  
XX (DIVE-) DIVERSA CORP.

XX PI  
XX Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;

XX DR  
XX WPI; 2004-121569/12.

XX DR  
XX P-PSDB; AD164287.

PT Novel isolated or recombinant polypeptide having nitrilase activity,  
PT useful in production of food additives.

PS Claim 1; SEQ ID NO 33; 105bp; English.

CC The invention describes an isolated or recombinant polypeptide (I) comprising amino acids having a sequence at least 50 % identical to a sequence (S1) available in electronic form (EC) from the following web site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its variants, having one or more mutations at residue 55 Lys, Gly or Glu, at residue 60 glutamic acid, at residue 111 Ser, their combinations or fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid; producing a chiral alpha-hydroxy acid molecule or a chiral amino acid molecule; producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid derivative or an (R)-phenyl lactic acid derivative; modifying a molecule; and for identifying a modified compound. The inventive method is useful for monitoring or determining % enantiomeric excess or % diastereomeric excess. (I) is useful in the production of food additives and drug intermediates. This sequence encodes a nitrilase of the invention.

CC  
XX  
SQ Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 12; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTCCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60  
Db 1 ATGTTAAGTCCCGTGACGAGTATCGCGCGCGCGGTGACAGCGCGCGCATCTTTCTC 60  
QY 61 GATCTCGACCGCACAGTGCAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAG 120  
Db 61 GATCTCGACCGCACCGTGCAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAG 120  
QY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATCCCGCTATCCCTTTGGATATGAGCTG 180  
Db 121 GTGCGCTGATCGCGTTTCCGAAACCTGGATTTCCCGCTATCCCGCTGATCTGAGCTC 180  
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAACTCGCTGTGCGC 240  
Db 181 GCGTCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAACTCGCTGTGCGC 240  
QY 241 GGCAGCAAGCATGGCGAGCGCCCTGGCGGATGCGCGCGCGCGCGCATGATGCTGTC 300  
Db 241 GGCAGCAAGCATGGCGAGCGCGCTGGCGGATGCGCGCGCGCGCGCATGATGCTGCTC 300  
QY 301 GCGGCTATAGCGAGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCGCGCCC 360  
Db 301 GTCGGCTTACGAGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCGCGCCC 360  
QY 361 GATGCGCATCTGATCGCGCGCGCGCGCGCTCAAGCCTTACCATGCGGAGCGACCGTG 420  
Db 361 GAAGGCGAGCTCATCGCGCGCGCGCGCGCTCAAGCCTTACCATGCGGAGCGAGGTG 420  
QY 421 TTCCGCGAGGAGAGCGGAGCGCATCTCCCGGTGCACGATACCGCCATCGGGCGCTCGC 480  
Db 421 TTCCGCGAGGAGGAGCGGAGCGCATCTCCCGGTTCAGAGACGGCGGTGTCATCGGC 480  
QY 481 GCGCTCTGTTGCTGGAGACATCCAGCATTGTGAAATACGCCATGTACGCCCGCGAC 540  
Db 481 GCCCTCTGCTGTGGAGACATCCAGCGCTCTCGAAATACCGCATGTATGCGGCAAC 540  
QY 541 GAACAGGTCCAGCTGCGGTGCGCGAGCTTACGCTTATCGCGGCGCATGCGCTATGCG 600  
Db 541 GAACAGGTGATGTGCGCTGCGCGGTGCTTACGCTTATCGCGGCGCATGCGCTATGCG 600  
QY 601 CTCGAGCGGAGGTCAATACCGCGCAAGCAGATCTACGCGGTGAGGCGGCTGCTAC 660  
Db 601 CTCGAGCGGAGGTGAACACCGCGCGAGCCAGGTCTACGCGGTGAGGCGGCTGCTAC 660  
QY 661 GTGCTGGGTGCTGCGGACCGTTTCGCCGAGATGATCAAGGTATTGTGATACGCCCC 720  
Db 661 GTGCTGGCTCTGTCTGTGTGACACCCGAGATCCTGAAGGTGCTGATCGACACGCCCC 720  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGTTTGGCCATGATTTTCGGGCGCGACGGC 780

Db 721 GACAAGAGCCGTTGCTGCTCGCGCGCGGGGGTCTCGATGATCTTCGCCCCGACGGC 780  
QY 781 CGCGCCTGGCCGAGCGCGCTCCCGAGACCGAAGAGGAGTCTGTGCGCGATATCGAC 840  
Db 781 CGCGCCTGCGCCAGCGCGCTGCGAGAGCCGAAGAGGGGCTGTCACGCGGAGATGAT 840  
QY 841 CTCGCGATGATCGCGTTGGCCAGCGCGCGCGCGCATCCGCGGGCCACTATTACGGCCC 900  
Db 841 CTCGCGCGATCGCGCTCGCCAGAGCGCGCGCGCATCCGCGGGCCATTACGCGCGCCC 900  
QY 901 GACGTAAAGCGGCTGCTGTGATCGAGTCCGCGCCCAACGCGTGTACGCTTGATGCC 960  
Db 901 GACGTAAAGCGGTTGTTGTGAACCGCGCGCGCGCGCGGTGGAAGCGTGGTCCG 960  
QY 961 GCATTGAAACCGCAAAACGAGACAAAGGCGGACGCGCGCGCGCTGCGGTGAGCGG 1018  
Db 961 CGCTTCAGGTGCTGTCAGAGCGAGCGAGCGCGCGCGCGCCCAACGAGCGGCGGAGCGG 1018

RESULT 12  
ADC23944  
ID ADC23944 standard; DNA; 1062 BP.  
XX  
AC ADC23944;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA sequence (SeqID 211) encoding a nitrilase enzyme.  
XX  
KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KM enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO200300840-A2.  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR P-PSDB; ADC23945.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.  
XX  
PS Claim 1; SEQ ID NO 211; 560pp; English.  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydrins into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polynucleotide is a DNA

CC sequence that encodes a nitrilase enzyme of the invention.

XX Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

SQL

Query Match 46.7%; Score 486; DB 10; Length 1062;

Best Local Similarity 67.8%; Pred. No. 2.3e-87;

Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

QY 17 CGAAGTATCGCGCGCGCGGTGACAGCGCGCGCGGTGTTCTCTGATCTCGACCGCACAG 76  
DB 17 CGAATATCAAAAGTGGCCCGCGGTGACAGCGCGCGCGGTGTTCTCGACCTGAGCGCCTCGG 76  
QY 77 TCGAGAAAAGCGATCGCGCTGATCGAGCAGCGCGCAAGCAGACAGCTGCGCTGATCGCAT 136  
DB 77 TCGAAAAGCGCGTCCGTTTCATCGACGAGCGCGCGCGCGCGCGCGCTCATCGCCT 136  
QY 137 TCCCAAGACTTGGATTCCCGGCTATCCCTTTTGGATATGCTGGCGCGCGCGCTTGGG 196  
DB 137 TTCCGAGAGCTGGATACCGGCTTACCCCTGGTGGATCTGGCTAGGCGCGCGCGCTGGG 196  
QY 197 GCAT-----GCGCTTCTGTCAGCGCTATTTCGAGAATTGCTCGTGGCGCGGACGACAGC 250  
DB 197 CTATCATGCGCGGCTTCTGCTCGCGCTATTTCGACAACTCGCTCAGCTACGACAGCCCGC 256  
QY 251 AGTGCAGAGCGCTGCGCGGATCGCGCGCGCGCGCGCGCATGATGCTGGCGCGCTATA 310  
DB 257 AGGCCGAGAAAGCTCCGCGCGCGCGCGCGCAAGCAAGATGGTGGTGGTCTCGCCTCT 316  
QY 311 GCGAGCGCGCGCGCGCGCGCGCTCTTATATGGGCGAGCGCATCTTGGCGCGCGCGATGCGCATC 370  
DB 317 CCGAGCGCGAGCGCGCGCGCGCGCGCTTTCATTCGCGCATGATCATCGGCCCGGACGCGGAAA 376  
QY 371 TGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCATCGAGCGCAAGCGTGTTCGCGAGG 430  
DB 377 CCATCGCGCAAGCGCGCGCAAGCTCAAGCGCAAGCGCGAGCGGACCGTGTTCGCGAAG 436  
QY 431 GAGACGCGAGCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGGCGCGCTCTGT 490  
DB 437 GCGACGCTCGCATCTTGGGTGACAGACTTGATGTGGCGCGCTCGGCGCGCTGTCT 496  
QY 491 GCTGGAGCATTCAGCCATGTGCAAAATACGCCATGTACCGCGCGGACGACAGAGTTC 550  
DB 497 GCTGGGAACAAGCTGACCGCGCTGTCAAAATACGCCATGTATGCGGAGAAAGAGTGC 556  
QY 551 ACGTCGCGTGTGCGCGCGGCTTCAAGCTTATCGCGCGCATGCGCTATGCGCTCGACCGG 610  
DB 557 ATGTGCGGCGCTGCGCGGAGCTTTCGCTTTACGATCCGTTGCGGACGCGCTCGCGCGG 616  
QY 611 AGTCAATACCGCGCGCAAGCTCAAGCTTACGCGGTGAGGCGCGCTGCTACGTGTGCGCT 670  
DB 617 AAGTGAACAATGCGCGGAGCAAAATCTATGCGGTGAGGCGCTGTCTGTCTCATGCGCG 676  
QY 671 CGTGGCGAGCGGTTTTCGCGGAGATGATCAAGGATATGTTGATGATACGCCGACAGAGAGA 730  
DB 677 CGTGGCGAGCGGTTTTCGAGGCGATGATGACGAACCTTGCATTAACGCCGAGAGAGCATC 736  
QY 731 TGTTCCTCAAGGCGCGCGCGGTTTTCATGATTTTCGCGCGGACGCGCGCGCGCTGG 790  
DB 737 AGTTCCTGATGCGCGCGCGCGGCTTTCGATTTACGCGCGCGCGCGCGCGCGCTCG 796  
QY 791 CCGAGCGCGCTCCCGGAGACCGAAGGGAAGTGTGCTGCTGCGCGGATATGACCTCGGACATGA 850  
DB 797 CCGCGCGCGCTGCGCGCGCGGACGAAAGGCTTGTCTACGCGCGCATCGATCTCGGAGTGA 856  
QY 851 TCGCGTTGGCCCAAGGCGCGCGCGGATTCGCGCGGCGCACTATTCACGCGCGAGCTAACGC 910  
DB 857 TTTCCGTTGCCAAAGCGGCGAGCGATTCGCGCGCGGCGCATTAAGCAGCGCGCGAGCTACCC 916  
QY 911 GGTGCTGCTGATCGAGCTCCGCGCGCAAGCGCGTGTGCTACGCTTGATGCGCATTCGAAC 970  
DB 917 GGTCTTCTGTTCAACAATCGCGCGCTGCTGATGCGGTGAGACCATGCGCTTGCAGATGATG 976  
QY 971 CGCAAAAAGAGGAGCAAGGCGGCGCGCGCGCTGCGCGCTGCGGTGGCGGAAAAGCGCGCGCG 1030

DB 977 CGAGACCAAGCGGAGGACCGGCTTAAGCCCGGAACCCAAAGGACCGAAGCTGGCGCGCT 1036  
QY 1031 CCGCGCAG 1038  
DB 1037 TCGCGCGG 1044

RESULT 13

ADH36045

ID ADH36045 standard; DNA; 1062 BP.

AC ADH36045;

DT 11-MAR-2004 (first entry)

DE Chemical process monitoring-related nitrilase gene sequence SegID211.

KW chemical process monitoring; biochemical process monitoring; cyanide;

KW high throughput system; gene; ds.

OS Unidentified.

PN WO2003098187-A2.

PD 27-NOV-2003.

PF 15-MAY-2003; 2003WO-US015639.

PR 15-MAY-2002; 2002US-0380737P.

PA (DIVE-) DIVERSA CORP.

PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;

PI Mcquaid J, Stege J;

DR MPI; 2004-142708/14.

DR P-PSDB; ADH36046.

PT Monitoring a chemical or biochemical process comprising providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.

XX Claim 74; SEQ ID NO 211; 277bp; English.

XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a gene which encodes a nitrilase enzyme which can be used in the  
CC method of the invention.

SQL Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

Query Match 46.7%; Score 486; DB 12; Length 1062;

Best Local Similarity 67.8%; Pred. No. 2.3e-87;

Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

QY 17 CGAAGTATCGCGCGCGCGGTGACAGCGCGCGCGGTGTTCTCTGATCTCGACCGCACAG 76  
DB 17 CGAATATCAAAAGTGGCCCGCGGTGACAGCGCGCGCGGTGTTCTCGACCTGAGCGCCTCGG 76  
QY 77 TCGAGAAAAGCGATCGCGCTGATCGAGCAGCGCGCGCAAGCAGAGCTGCGCTGATGCAAT 136  
DB 77 TCGAAAAGCGCGTCCGTTTCATCGACGAAAGCGCGCGCGCGCGCGCTTCAATGCTT 136  
QY 137 TCCCAAGACTTGGATTCCCGGCTATCCCTTTTGGATATGCTGGCGCGCGCGCTTGGG 196



Db 137 TTCCGAGACCTGATACCCGGTTACCCCTGGTGATCTGGCTAGGCGCGCGCCTGGG 196  
QY 197 GCAT-----GCGCTTCGTCCAGCGCTATTTCAGAAATTCCGCTCGTCCGCGCAGCAAGC 250  
Db 197 CTATCATGCGCGGCTTCGTCTCGCGCTATTTCGACACTCGCTCAGCTACGACAGCCCGC 256  
QY 251 AGTGGCAGGCCCTGGCGCGATGCGGCCCGCCCGCCACGGCATGCATGTCGTGGCCGGCTATA 310  
Db 257 AGGCCGAGAAGCTCCGCGCGCCCGCCGCAAGCGCAACAAGATGGTGTGTGCTCGGCCCTCT 316  
QY 311 GCGAGCGCGCGCGCGCAGCCTCTATATGGCCAGCGCATCTTCGGCCCCCGATGGCGATC 370  
Db 317 CCGAGCGCAGCGCGCGCAGCCTTTACATCGCGCAATGATCATCGGCCCGGACGGCGAAA 376  
QY 371 TGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGGTTCGGCGAGG 430  
Db 377 CCATCGCCAAAGCGCGCGCAAGCTCAAGCCGACCGCGAGCGGACCGGTTCGGCGAAG 436  
QY 431 GAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGCGGCTCTGTT 490  
Db 437 GCGACGGCTCGCATCTTCCGGTGACAGAGCTTGATGTGGCCGGCTCGGCGGCTGTGCT 496  
QY 491 GCTGGAGCACAATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGCAACAGGTCC 550  
Db 497 GCTGGAAACACTGCAAGCCGCTGTCCAAATACGCCATGTATGCGCAGAACGAACAGGTGC 556  
QY 551 ACGTCGCGTGTGGCGAGCTTCAGCCTCTATCGCGGCATGGCCTATGCGTCCGACCGG 610  
Db 557 ATGTCCGCGCCTGGCGGAGCTTTTCGTTTACGATCCGTTCCGCGACGCGCTCGGCGCGG 616  
QY 611 AGTCAATACCGCGCGCAGCCAGATCTACGCGGTGAGGGCGCGCTACTAGCTGCTGGCGT 670  
Db 617 AAGTAACAATGCGCGGAGCAAAATCTATGCGGTGAGGGCTCGTGTTCGTCATCGCGC 676  
QY 671 CGTGGCGACCGTTTCGCCGAGATGATCAAGTATTGCTGATACGCCCGCAAGAGAGA 730  
Db 677 CGTGGCGACCGTTTCGAGGCGATGATCGACGAACCTGCGGATACGCCCGGAGAAAGCATC 736  
QY 731 TGTTCCTCAAGCGCGCGCGCGGTTTGGCATGATTTTCGGGCCCGACCGCGCGCCCTGG 790  
Db 737 AGTTCCTGATGCGCGCGCGCGGCTTTGCGGTGATTTACGGCCCCGACGGCGCGCCTCG 796  
QY 791 CCGAGCGGCTCCCGGAGACCGAAGAGGGAGCTGTGCTGCGCGATATCGACTCGGCATGA 850  
Db 797 CGGCGCGGCTGCGCGCGCAAGGAAGGCTTGTCTACGCCGACATGATCTCGGAGTGA 856  
QY 851 TCGCGTTGGCAAGCGCGCGCGCGATCCGCGCGGCCACTATTCAAGGCCCGCATGAAGC 910  
Db 857 TTTCGTTGGCAAGCGCGCGAGCCGATCCGCGCGGCATTTATGACAGCCCGCAAGTCAACC 916  
QY 911 GGCTGCTGTGATCGACGTCCGGCCCAACGCGTGTACGCTTGATGCCGATTCGAAC 970  
Db 917 GGCTTCTGTCAACAATCGGCCCTGGGTATCGGTCGAGACCATGCGCTTGCGATCGATG 976  
QY 971 CGCAAAACGAGACAAAGGCGAGCGCGCGCGCTGCGCGTGGTGCGGAAGCGCGCGCG 1030  
Db 977 CGGAGACCAAGGCGGAGACACCGGCTAAGCCGGAACCCAAAGCACCGAAGCTGGCGCGT 1036  
QY 1031 CCGCGCAG 1038  
Db 1037 TCGCGCCG 1044

RESULT 14  
ADG93746  
ID ADG93746 standard; DNA; 1062 BP.  
XX  
AC ADG93746;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitrilase enzyme gene sequence SeqID211.  
XX

KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; gene; ds.  
OS Unidentified.  
XX  
PN WO2003097810-A2.  
XX  
XX 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
XX (DIVE-) DIVERSA CORP.  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-090637/09.  
DR P-PSDB; ADG93747.  
XX  
PT New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
PS Claim 44; SEQ ID NO 211; 295pp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

Query Match 46.7%; Score 486; DB 12; Length 1062;  
Best Local Similarity 67.8%; Pred. No. 2.3e-87;  
Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;  
QY 17 CGAAGTATCGGCGCGCGGCTGATCGAGCGCGCGCGGTGTTCTCGATTCGACCGCACAG 76  
Db 17 CGAATACAAAGTGGCGCGCGTGCAGGACGCGCGCTTTCTCGACCTGGAACGCTCGG 76  
QY 77 TCGAAGAGCGATCGGCTGATCGAGCAGCGCGCCAGCAGAGCAGTGGCCTGATCGCAT 136  
Db 77 TCGAAAAGCGGCTCCGTTTCATCGACGAGCCGCGCGCGCGCGCTTCATCGCCT 136  
QY 137 TCCAGAGACTTGATTTCCGCGCTATCCCTTTGGATATGCTGGCGCGCGCTTGGG 196  
Db 137 TTCCGAGACCTGATACCCGTTACCCCTGTGATCTGGCTAGGCGCGCGCTGGG 196  
QY 197 GCAT-----GCGCTTCGTCCAGCGCTATTTCGAGAATTGCTCGTGGCGCGCAGCAAGC 250  
Db 197 CTATCATGCGCGGCTTCGTCTGCGGCTATTTCGACAACTCGCTCAGCTACGACAGCCCGC 256  
QY 251 AGTGGCAGGCCCTGGCGGATGCGGCCCGCCGACGGCATGATGTCGTGGCCGGCTATA 310  
Db 257 AGGCCGAGAAGCTCCGCGCGCGCGCCGCAAGCGCAACAAGATGGTGTGTGCTCGGCCCTCT 316  
QY 311 GCGAGCGCGCGCGCGCAGCCTCTATATGGCCAGGCGATCTTCGCGCCGATGGCGATC 370  
Db 317 CCGAGCGCGAGCGCGCAGCCTTTACATCGCGCAATGATCATCGGCCCGGACGGCGAAA 376  
QY 371 TGATCGCGCGCGCGCGCAAGCTCAAGCCTAACCATGCGAGCGCAGCGTGTTCGGCGAGG 430  
Db 377 CCATCGCCAAAGCGCGCAAGCTCAAGCCGACCCAGCGGAGCGGACGTGTTCGGCGAAG 436  
QY 431 GAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGCGCGCTCGGCGGCTGTGTT 490  
Db 437 GCGACGGCTCGCATCTTCCGGTGACAGAGCTTGATGTGGCCGGCTCGGCGGCTGTGCT 496



QY	491	GCTGGGAGCACATCCAGCCATTGTCGAATACGCCATGTACCCGCCGACGAACAGTCC	550
Db	497	GCTGGGAACACCTGCAGCCGCTGTCCAATACGCCATGTATGCGACAAGAACAGGTGC	556
QY	551	ACGTCCGCTGCTGGCCGAGCTTCAGCCTCTATCGCGGCATGCGCTATGCGCTCGACCG	610
Db	557	ATGTCCGGCCTGGCCGAGCTTTTCGCTTACGATCCGTTCCGCGACGCGCTCGCGCG	616
QY	611	AGCTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTACGTGTCGCT	670
Db	617	AAGTGAACATGCGGCGAGCAAAATCTATGCGGTGAGGCTGTGTTTCGTCATCGCG	676
QY	671	CGTCCGCGACCGTTTCGCGCGAGATGATCAAGTATTGTGATACGCCCGACAAGAGA	730
Db	677	CGTCCGCGACCGTTTCGAGCGCATGATCGACGAACCTCTGCGATACGCCGAGAAGCATC	736
QY	731	TGTTCTCAAGCCGCGCGCGTTTGCATGATTTTCGGCCCGACGCGCGCGCTGG	790
Db	737	AGTTCCTGCATGCGCGCGCGCTTTCGCGTATTTACGGCCCCGACGCGCGCGCTCG	796
QY	791	CCGAGCCGCTCCCGGAGACCGAAGGGACTGCTGTCGCCGATATCGACCTCGGCA TGA	850
Db	797	CGGCGCGCTGCCGCCGACGAAGGAGCTTGCTCTACGCCGACATCGATCTCGGGATGA	856
QY	851	TCCGCTTGCCCAAGCGCGCGCGCATCCGCGGGGCCACTATTCACGGCCGACGTACGC	910
Db	857	TTTCCGTTGCCAAGCGGCGACGCCGATCCGCGCGGCATTATGCACGCCCCCGACGTACCC	916
QY	911	GGCTGCTGTGATCGACGTCCGCCCCAACGCGTCTGTACGCTTGATGCCGATTGCAAC	970
Db	917	GGCTTCTGTTCAACAATCGGCTTGCGTATCGGGTGCAGACCATGCGCTTGCCGATCGATG	976
QY	971	CGCAAAACGAGGACAAAGGCCGACGCCGCCGCTGCGCGGTGGCGGAAAGCGCGCGCG	1030
Db	977	CGGAGACCAAGCGCGGAGGACCTGAGCTTAAGCCGGAAACCAAGGCCACCGAACGTGGCGCGT	1036
QY	1031	CCGCGCGAG 1038	
Db	1037	TCCGCGCCG 1044	

RESULT	15
AD162343	
ID	AD162343 standard; DNA; 1062 BP.
XX	
AC	AD162343;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	DNA encoding nitrate reductase polypeptide #106.
XX	
KW	Acetoacetic; (R)-ethyl 4-cyano-3-hydroxybutyrate;
KW	(R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;
KW	3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;
KW	4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;
KW	mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;
XX	antihypertensive; gene; ds.
XX	
OS	Unidentified.
XX	
PN	WO2003106415-A2.
XX	
PD	24-DEC-2003.
XX	
PF	13-JUN-2003; 2003WO-US018840.
XX	
PR	13-JUN-2002; 2002US-0389317P.
XX	
PA	28-JUN-2002; 2002US-0392944P.
XX	
PI	(DIVE-) DIVERSA CORP.
XX	
PI	Burk M, Desantis G, Morgan B, Zhu Z;

XX WP1; 2004-090821/09.  
DR P-PSDB; ADI62344.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 46; SEQ ID NO 211; 253bp; English.  
XX  
XX The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence encodes a nitrilase polypeptide obtained from an environmental  
CC sample.  
XX  
SQ Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

Query Match	46.7%	Score 486	DB 12	Length 1062
Best Local Similarity	67.8%	Pred. No. 2.3e-87		
Matches 697	Conservative 0	Mismatches 325	Indels 6	Gaps 1
QY 17	CGAAGTATCGCGCGCGCGGTGCAGAGCCGCCCGGTCTCGATCTGACCCGCACAG			76
Db 17	CGAATATACAAAGTGGCCGCGGTGCAGGACGCCCGCTTTCTCGACCTGAGCGCTCGG			76
QY 77	TCGAGAAAGCGATCGGCGCTGATCGAGCAGCGCGCCAAAGAGACGTGCGCTGATCGCAT			136
Db 77	TCGAAAAGGCGTCCGTTTCATCGACGAAGCCGCGCGCGCGCGCTCATCGCCT			136
QY 137	TCCCAGAGACTTGATTCGCCGCTATCCCTTTGGATATGCGTGGCGCGCGCTTGG			196
Db 137	TTCCGAGACCTGATACCCGGTTACCCCTGTGATCTGGCTAGCGCGCGCGCTGGG			196
QY 197	GCAT-----GCGCTTCGTCCAGCGCTATTTGAGAAATTCGCTCGTGCAGCGCAAGC			250
Db 197	CTATCATGCGCGCTTCGTCTCGCGCTATTTCAGACATCGCTCAGCTACGACAGCCCGC			256
QY 251	AGTGCAGGCGCTTGGCGGATGCGGCCCGCGCACGGCATGTCGTGCGCGCTATA			310
Db 257	AGGCCGAGAAAGCTCCGCGCGCGCGCCGACGCAACAGATGGTGTGTGCTCGGCTCT			316
QY 311	GCGAGCGCGCGCGCGCGACCTCTATATGGGCCAGGCGATCTTCGGCCCCGATGGCGATC			370
Db 317	CCGAGCGCGACGCGCGCGCGACCTTTACATCGCGCAATGATCATCGGCCCGGACGCGCAA			376
QY 371	TGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATCGCGAGCGCACCGTGTTCGCGGAG			430
Db 377	CCATCGCCAAAGCGCGCGCAAGCTCAAGCCGACCCACGCGGAGCGGACCGTTCGCGCAAG			436
QY 431	GAGACGGCAGCCATCTCGCGGTGCAAGATACCGGCATCGGCGCGCTCGGCGCGCTTGT			490
Db 437	GCGACGGCTCGCATCTTGCGGTGCAAGAGCTTGATGTGGCGCGCGCTCGGCGCGCTGTGCT			496
QY 491	GCTGGGAGCATCCAGCCATTGTGAAATACGCCCATGTACGCCCGGACGAACAGGTCC			550
Db 497	GCTGGAAACACTGACCGCGCTGTCCAAATACGCCCATGTATGCGCAGAAACGAACAGGTGC			556
QY 551	ACGTGCGCTGTGGCGGAGCTTCAGCCTCTATCGCGCATGGCGCTATGCGCTCGGACCGG			610
Db 557	ATGTGCGGCGCTGCGGAGCTTTTCGCTTACGATCCGTTGCGGCACGCGCTCGGCGCGG			616

```

QY 611 AGTCAATACCGCCGCAAGCCAGATCTACCGGTCGAGGGCGGCTGCTACGTGCTGGCGT 670
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 AAGTAACAATGCGCGAGCAAAATCTATGCGGTGAGGGCTCGTGTTCGTATCGCGC 676

QY 671 CGTGGCGACCGTTTCGCCGAGATGATCAAGTATTTGGTATACGCCCGACAAGAGA 730
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 CGTGGCGACCGTTTCGCAAGCGATGATCGACGAACCTGCGATACGCCGAGAAAGCATC 736

QY 731 TGTTCCTCAAGCGCGCGCGGTTTTCGATGATTTTCGGGCCGACGCGCGCCCTGG 790
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 AGTTCCTGCATGCGCGCGCGGCTTTGCGGTGATTTACGGCCCCGACGCGCGCCGCTCG 796

QY 791 CCGAGCCGCTCCCGGAGACCGAAGAGGACTGCTGTGCGCGATATCGACCTCGGCATGA 850
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 797 CCGCGCCGCTGCGCGCGCGACCAAGGAAGGCTTGCTTACGCCGACATCGATCTGGGATGA 856

QY 851 TCGCGTTGCCAAGCGCGCGCGCGATCCGCGCGGCCACTATTCACGGCCCGACGTAAAGC 910
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 857 TTTCCGTTGCCAAGCGCGCGCGATCCGCGCGGCATTTATGCACGCCCGACGTCAACC 916

QY 911 GGCTGCTGTGATCGACGTCCGCGCCCAACGCGCTGCTCAAGCTTGATGCCGATTCGAAC 970
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 917 GGCTTCTGTCAACATCGGCTTGATCGGTCGAGACCATGAGCGTTGCCGATCGATG 976

QY 971 CGCAAAACGAGACAGGGCGACGCGCCCGCGCTGCGCTGTGTGCGGAAAGCGCCCGCG 1030
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 977 CGGAGACCAAGCGCGAGACACCGGCTTAAGCCGGAACCAAGGACCGAACGTGCGCGCGT 1036

QY 1031 CCGCGCAG 1038
    ||||| |||||
Db 1037 TCGCGCCG 1044

```

Search completed: April 26, 2006, 22:54:28  
 Job time : 721.835 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:54:39 ; Search time 214.785 Seconds  
(without alignments)  
8615.305 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1041  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCITUS COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	343.2	33.0	1035	2	US-08-690-493-2
2	343.2	33.0	1200	2	US-08-690-493-3
3	297.4	28.6	1071	3	US-09-806-876A-1
4	283.8	27.3	1110	3	US-09-823-373-4
5	283.8	27.3	1110	3	US-09-823-373-13
6	283.8	27.3	1776	3	US-09-823-373-15
7	202	19.4	1110	3	US-09-823-373-16
8	132.2	12.7	385	3	US-09-823-373-3
9	109.6	10.5	1194	2	US-08-447-702-4
10	108	10.4	1194	2	US-08-465-615-4
11	57	5.5	1995	3	US-09-902-540-4902
12	57	5.5	28493	3	US-09-902-540-1241
13	55	5.3	1728	3	US-09-758-759-126
14	55	5.3	109519	3	US-09-758-759-1
15	54.6	5.2	5877	3	US-10-152-886-94
16	51.8	5.0	5857	3	US-09-477-962-1
17	51.6	5.0	507	3	US-09-902-540-4718
18	51.6	5.0	5121	3	US-09-902-540-704
19	51.6	5.0	28762	3	US-09-902-540-1232
20	51.4	4.9	2124	3	US-09-902-540-7586
21	51.4	4.9	7704	3	US-09-902-540-743
22	50.8	4.9	105413	3	US-10-427-923-3
23	50.8	4.9	112219	3	US-09-949-016-12453
24	50.8	4.9	112222	3	US-09-949-016-14324

C	25	50.4	4.8	109519	3	US-09-758-759-1	Sequence 1, Appli
	26	49.8	4.8	30001	2	US-08-125-468-1	Sequence 1, Appli
	27	49.8	4.8	30001	2	US-08-474-933-1	Sequence 1, Appli
	28	49.4	4.7	1377	3	US-09-252-991A-10413	Sequence 10413, A
	29	49.4	4.7	1383	3	US-09-252-991A-10306	Sequence 10306, A
C	30	49.4	4.7	1410	3	US-09-252-991A-10749	Sequence 10749, A
	31	48.8	4.7	1494	3	US-09-902-540-2363	Sequence 2363, Ap
C	32	48.8	4.7	15447	3	US-09-902-540-1100	Sequence 1100, Ap
	33	48.6	4.7	1098	3	US-09-252-991A-14407	Sequence 14407, A
C	34	48.6	4.7	1566	3	US-09-252-991A-14264	Sequence 14264, A
	35	48.6	4.7	1908	3	US-09-252-991A-14353	Sequence 14353, A
	36	48.4	4.6	699	3	US-09-252-991A-4020	Sequence 4020, Ap
	37	48.2	4.6	1377	3	US-09-377-557-17	Sequence 17, Appl
	38	47.6	4.6	939	3	US-09-902-540-6060	Sequence 6060, Ap
C	39	47.6	4.6	941	3	US-09-902-540-252	Sequence 252, App
	40	47.6	4.6	2838	3	US-09-252-991A-14128	Sequence 14128, A
	41	47.6	4.6	3897	3	US-09-252-991A-14008	Sequence 14008, A
	42	47.4	4.6	3957	3	US-10-237-551-193	Sequence 193, App
	43	47.4	4.6	23673	3	US-09-773-816-1	Sequence 1, Appli
	44	47.4	4.6	154746	3	US-09-827-688-8	Sequence 8, Appli
C	45	47.4	4.6	154746	3	US-09-827-688-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-2  
; Sequence 2, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000e1 Nitrilase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1035 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Gordona terrae  
; STRAIN: MA-1  
; CELL TYPE: unicellular organism  
US-08-690-493-2  
Query Match 33.0%; Score 343.2; DB 2; Length 1035;



```
Best Local Similarity 61.8%; Pred. No. 1.5e-67;
Matches 546; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

QY 31 GCGGCGGTGCAGGCGCGCGCGGTGTCTCTCGATCTGCACCGCACAGTGCAGAAAGCGATC 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 34 GCGGTGACCCAGCGCGAAACCGGTGTGTCTGACTCTCGGCCACCGCTGCACAAGACCATT 93

QY 91 GGCCTGATCGAGACGCGCGCGCAAGACGACGCGCGCTGATCGCATTCGAGAGACTTGG 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 94 GCGCTGCTGAGAGAGCGCTCCCGCGCGCGCATCTGATCGCGTTCCCGAGAGACTTG 153

QY 151 ATTCCCGCTATCCCTTTTGGATATGCTGGCGCGCGCGCTTGGGGCATGCGCTTCTC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 ATACCGGGTACCGGTGTCTCTGTGCTTGAATCGGTGGCTTGCAGAGCAGTACTTC 213

QY 211 CAGCGCTATTCGAGAAATTCGCTCGTGGCGCGCGAGCAAGCATGGCAGGCCCTGGCGAT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 214 ATCCGATATCCGAGAACTCGCTCGATCTCGACGGCAGCGAGTTCCGCGGATCAGGAA 273

QY 271 GCGGCGCGCGCGCAGCGCATGCATGTCTGCGCGCGCTATAGCGAGCGCGCGCGGAGC 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 GCGGACAGAAAGAACGACATCGCGATCACCATGGGATTTAGTAGCGCGCTCATGTTG 333

QY 331 CTCTATATGGGCGCAGCGCATCTTCCGCGCGCGATGGCGATCTGATCGCGCGCGCGAG 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 334 CTGTACATGGGCGCAGCGCGTCTCATCGAGCGGTGACGGGGTCTGCTACGACACGCCGCAA 393

QY 391 CTCAGCGCTACCCATGCGGAGCGCAGCGCGTGTTCGGCGAGGAGAGACGCGACCATCTCG 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 394 CTGAAGCCGACCCACGTCGAGCGGACCCCTGTTCGGTGAAGGTGATGTTCCGATCTGTC 453

QY 451 GTGACGATACCGCCATCGGCGCGCTCGCGCGCTCTGTGCTGGAGACATCCAGCCA 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 454 GTGACCCAGACCACTCTCGGCGGAGTGGGTGCTGTGTTGGAAATCTGCAGCCG 513

QY 511 TTGTCGAAATACGCCATGTACGCCCGCGCAAGAACAGTCCACGTCGCTGCTGGCGGAGC 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 TTGACCAAGTACGCCATGTACTCGCAGCAGACAGATTCACATCGCGCATGGCGCCAGC 573

QY 571 TTCAGCCTTATCGCGCGCATGCGCTATGCGCTCGGACCGAGGTCAATACCGCGCAAGC 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 TTCTGATCTTCCGGCGCGGTGTATGCGCTCGGCGCGAGGTCAACACCGCGCTCT 633

QY 631 CAGATCTACGCGGTGAGGCGCGCTGTAAGTGTGCTGCGCTCGTGGCGGACCGTTTCGCG 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 634 CAGCAATACGCGGTAGAGGCGAGACCTTACGTTCTCGCTCCATGCGCGGTATCGCGAT 693

QY 691 GAGATGATCAAGTATTGTTGATACGCGCGCAAGGAGATGTTCTCAAGCGCGCGC 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 694 GCAGTTGGAGGCGTTTCCGATACCGAGAGAGACGACGCTCATCCACAAGAGAGC 753

QY 751 GGTTTTGCATGATTTTTCGGGCGCGACGCGCGCGCTTGGCGGACCGCTCCCGAGACC 810
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 754 GGATATGCGCGTATCTACGTCGCCGACGCTGTTCACTCGCGGAACCGCTCGCGCCAAT 813

QY 811 GAAGAGGACTGCTGCTGCGCGCATATCGACTCGGCATGATCGCGTTGGCCAGAGCGCG 870
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 814 GACGAGGGAATCTGTACGCGGACATCATCTGTCTGCGATTCGCGCCGCAAGAACCCG 873

QY 871 GCGGATCCGCGCGCGCATATTACGCGCGCGACGTAACCGCGCT 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 874 GCGGACCGCGTGGGCACTACTCGCGTCCGACGTACTGCGTCT 917

RESULT 2
US-08-690-493-3
; Sequence 3, Application us/08690493
; Patent No. 5872000
; GENERAL INFORMATION:
; APPLICANT: Yu, Fujio
; TITLE OF INVENTION: No. 5872000e1 Nitrilase Gene
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
;
```

```
ADDRESSEE: Steinberg, Raskin & Davidson, P.C.
STREET: 1140 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MS-DOS Editor
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/690,493
FILING DATE: 31 JUL-1996

PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 213061/1995
FILING DATE: 31-JUL-1995
ATTORNEY/AGENT INFORMATION:
NAME: Davidson, Clifford M
REGISTRATION NUMBER: 32,728
REFERENCE/DOCKET NUMBER: 3821005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 768-3800
TELEFAX: (212) 382-2124
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1200 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Gordona terrae
STRAIN: MA-1
CELL TYPE: unicellular organism

US-08-690-493-3

Query Match 33.0%; Score 343.2; DB 2; Length 1200;
Best Local Similarity 61.8%; Pred. No. 1.6e-67;
Matches 546; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

QY 31 GCGGCGGTGCAGGCGCGCGGTGTCTCTCGATCTGCACCGCACAGTGCAGAAAGCGATC 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 GCGGTGACCCAGCGCGAACCAGTGTGTTGCACCTCTCGGCCACCGTCGACAAGACCATT 187

QY 91 GGCCTGATCGAGCAGCGCGCGCAAGCAGACGTCGCGCTGATTCGATTCGCCAGAGACTTGG 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 GCGCTGTCGAAGAGGCGCTCCGCGCGCGCGCGCGATGATCGCGTCCCGAGACTTG 247

QY 151 ATTCGCGCTATCCCTTTTGGATATGCTGCGCGCGCGCGCTTGGGCGATCGCGCTTCTC 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 248 ATACCGGGTACCCGTGTCTGTGCTTGAATCGCTTGAATCGCGGAGGAGTTGCGCGGATCAGGAA 307

QY 211 CAGCGCTATTTCGAGAAATTCGCTGTCGCGCGGAGCAAGCATGGAGCGCGCTGGCGGAT 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 308 ATCCGATATCCGAGAACTGCTCGATCTGACCGGAGGAGTTGCGCGGATCAGGAA 367

QY 271 GCGGCGCGCGCGCAGCGCATGCATGCTGTCGCGCGGCTATAGCGAGCGCGCGCGGAGC 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 GCGGACGAAAGAACGACATCGCGATCACCATGGGATTTAGTAGCGCGGTGATGTTG 427

QY 331 CTCTATATGGGCGCAGCGATCTTCCGCGCGCGCGCGCGATGATCGCGCGCGCGCAAG 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 CTGTACATGGGCGCAGCGGTCTCATGAGCGGTGACGGGTCGTCTACGACACGCGCAAA 487

QY 391 CTCAGCCTACCCATGCGGAGCGCACCGGTGTTGCGGAGGAGAGCGGACGACATCTCGCG 450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 CTGAAGCCGACCCACGTCGAGCGGACCGCTGTTGCGTGAAGGTGATGATTCGATCTGTC 547

QY 451 GTGACGATACCGCATCGGCGCGCTGCGCGCGCTGTGTTGCTGGAGACATCCAGCCA 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 GTGACCAAGACCACTCTCGGCGGAGTGGGTGCTGTGTTGGAACATCTGCAGCCG 607
```



```
Query Match      27.3%; Score 283.8; DB 3; Length 1110;
Best Local Similarity 57.0%; Pred. No. 2.8e-54;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

QY 19 AAGTATCGCGCGCGCGGTGCAGGCGCGCGCGGTCTCTCGATCTCGACCGCACAGTC 78
    |||||
    19 AAGTTCCTCGCGGCAACCGTTCAAGGCGAGAGCCGGTATGGCTCGACGCGACGCAACGATC 78
QY 79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGGCAAGCAGACGTCGCTGATCGCATTC 138
    |||||
    79 GACAAAGTCGATCGGCATCATCGAAGAGCTGCCAAAGGGCGCGAGTCGATCGCTTTC 138
QY 139 CCAGAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTGGGCGCGCGCTGGGGC 198
    |||||
    139 CCGGAAGTATTCATTCCGGGCTACCCCTATTTGGCGCTGGCTCGGCGACGTGAAGTACAGC 198
QY 199 ATGCGCTTCTCGACAGCGCTATTTGAGAAATTCGCTCGTCGCGCGCAGCAGCATGGCAG 258
    |||||
    199 CTAAAGCTTACTTCAAGCTATCAGAGAAATTCGTTGAGAGCTAGGTGACGACCGTATGCGT 258
QY 259 GCCCTGGCGAGTCGCGCGCGCGCGCGCGCGCATGTCGTGCGCGGCTATAGCGAGCGC 318
    |||||
    259 CGCTTCAGCTGGCGCGCGCGCGCGCAACAAATCGCACTCGTCATGGGCTATTCGAGCGCG 318
QY 319 GCGGCGCGCAGCCTCTATATGGCGCAGCGCATCTTCGCGCGCGCGATGGCGATCTGATCGCC 378
    |||||
    319 GAAGCCGAGTCGCGCTATCTGAGCCAGGTTCATCGACGAGCGCTGGCGAGATCGTTGCC 378
QY 379 GCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGGTTCGCGCGAGGAGACGCGC 438
    |||||
    379 AATCGCGCAAGCTGAAGCCACACACGTTAGCGTACGATCTACGCGCAAGGCAACGGA 438
QY 439 AGCCATCTCGCGGTGCAAGATACCGCCATCGGCGCGCTCGCGCGCTCTGTCTGGGAG 498
    |||||
    439 ACCGATTTCTCAAGCAGCACTTCGCGGTTCGACGCGCTCGGTGATGAACCTGCTGGAA 498
QY 499 CACATCCAGCCATTGTGCAAAATACGCCCATGTAAGCGCGCGCAGCAACAGTCCACGTCGCG 558
    |||||
    499 CATTTCCAAACCGCTCAGCAAGTTCATGATGTAAGCCTCGGTGAGCAGGTTCCACGTTGCA 558
QY 559 TCGTGCGCGAGCTTCAGCCTCTATCGCGGATGGCCTATGCGCTCGGACCGGAGGTCAT 618
    |||||
    559 TCGTGCGCGGAGTGTCCCTCTTCAGCGCGAGTGTTCCAACTGACATCGAAGGCCAAC 618
QY 619 ACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGCGCTGCTACGTCGTGCGTCGCGCG 678
    |||||
    619 GCGACGGTCAACCGCTCGTACGCAATGGAAGCCAAACCTTTGTGCTTGTGCTCGACGCGAG 678
QY 679 ACCGTTTCGCGGAGATGATCAAGGTATGTGTGATACGCCCGACAGAGATGTTCTTC 738
    |||||
    679 GTGATCGGAAGCTTACCGCATCGAAACGTTCTGCTCAACGACGAAACAGCGCACGCTGTG 738
QY 739 AAGCGCGCGCGGCTTTTGCCATGATTTTCGGGCGCGCGCGCGCGCGCTGGCCGAGCGC 798
    |||||
    739 CCGCAAGGATGTGGCTGGCGCGCATTTAACGGCCCGGATGGAAGCGAGCTTGCGAAGCCT 798
QY 799 CTCCCGGAGACCGAAGAGGACTGCTGTGCGCGGATATCGACCTCGCATGATCGCGTTG 858
    |||||
    799 CTGCGGGAAGATGCTGAGGGGATCTTGTATCGCAGAGATCGATCTGAGACAGATTCTGCTG 858
QY 859 GCCAAGGCGCGCGCGCGATCCGCGCGCACTATTACAGGCGCGCGACGTAACGCGGCTGCTG 918
    |||||
    859 GCGAAGGCTGAGCGCGATCCGGTCGGGCACTATTGCGGCGCTGACGTGCTGTGCTCGAG 918
QY 919 CTGATCGACG 929
    |||||
    919 TTCGACCCGCG 929
Db
```

```
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCobimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 13
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Acidovorax facilis
US-09-823-373-13
```

```
Query Match      27.3%; Score 283.8; DB 3; Length 1110;
Best Local Similarity 57.0%; Pred. No. 2.8e-54;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

QY 19 AAGTATCGCGCGCGCGGTGCAGGCGCGCGCGGTCTCTCGATCTCGACCGCACAGTC 78
    |||||
    19 AAGTTCCTCGCGGCAACCGTTCAAGGCGAGAGCCGGTATGGCTCGACGCGACGCAACGATC 78
QY 79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGACGTCGCGCTGATCGCATTC 138
    |||||
    79 GACAAAGTCGATCGGCATCATCGAAGAGCTGCCAAAGGGCGCGAGTCGATCGCTTTC 138
QY 139 CCAGAGACTTGATTCGCGGCTATCCCTTTGGATATGGCTGGGCGCGCGCTGATCGCATTC 198
    |||||
    139 CCGGAAGTATTCATTCCGGGCTACCCCTATTTGGCGCTGGCTCGGCGACGTGAAGTACAGC 198
QY 199 ATGCGCTTCTCGACAGCGCTATTCGAGAAATTCGCTCGTCGCGCGCGCAGCAAGCAGTGGCAG 258
    |||||
    199 CTAAAGCTTACTTCAAGCTATCAGAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGT 258
QY 259 GCCCTGGCGAGTCGCGCGCGCGCGCGCGCATGTCGTGCGCGCTATAGCGAGCGC 318
    |||||
    259 CGCTTCAGCTGGCGCGCGCGCGCAACAAATCGCACTGTCATGGGCTATTCGAGCGCG 318
QY 319 GCGGCGCGCAGCCTCTATATGGGCGCAGCGCATCTTCGCGCGCGCGATGGCGATCTGATCGCC 378
    |||||
    319 GAAGCCGAGTCGCGCTATCTGAGCCAGGTGTTCAATGACGAGCGCTGGCGAGATGTTGCC 378
QY 379 GCGGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGGTTCGCGGAGGAGACGCGC 438
    |||||
    379 AATCGCGCAAGCTGAAGCCACACACGTTGAGCGTACGATCTACGCGAAGGCAACGGA 438
QY 439 AGCCATCTCGCGGTGACGATACCGCCATCGGCGCGCTCGCGCGCTGTGCTGGGAG 498
    |||||
    439 ACCGATTTCTCAAGCAGCACTTCGCGTTCGACGCGCGCTGGATGAACTGCTGGAA 498
QY 499 CACATCCAGCCATTGTGCAAAATACGCCATGTACGCGCGGACGAAACAGTCCACGTCGCG 558
    |||||
    499 CATTTCCAAACCGCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558
QY 559 TCGTGCGCGAGCTTCAGCCTCTATCGCGGATGGCCTATGCGCTCGACCGGAGTCAAT 618
    |||||
    559 TCGTGCGCGGAGTGTCCCTCTTCAGCCGATGTTTCCAACTGAGCATCGAAGCCAAC 618
QY 619 ACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGCGCTACGTCGTGCGCTGCGCGC 678
    |||||
    619 GCGACGGTCAACCGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGCTCGACGCGAG 678
QY 679 ACCGTTTCGCGGAGATGATCAAGGTATGTGTGATACGCCCGACAGAGAGATGTTCTTC 738
    |||||
    679 GTGATCGGAAGCTTACCGCATCGAAACGTTCTGCTCAACGACGAAACAGCGCGCACGCTGTTG 738
Db
```



```
QY 739 AAGCGCGCGGTTTGGCATGATTTTCGGCCCGACGGCCGCGCCCTGGCGGAGCCG 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 739 CCGCAGGATGTGGCTGGCGCGCATTTACGGCCCGATGGAAGCGAGCTTGGAAAGCCT 798
QY 799 CTCCCGAGACCGAAGAGGAGTCTGCTGTCGCCGATATCGACCTCGGCGATGATCGCGTTG 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 799 CTGGCGGAAGATGCTGAGGGGATCTTGATCGCAGAGATCGATCTGAGCAGATTTCTGCTG 858
QY 859 GCCAAGCGCGCGCCGATCCGGCGGGCCACTATTCAAGCGCCGACGTAAAGCGGCTGCTG 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 859 GCGAAGGCTGAGCCGATCCGGCTGGGCACTATTCCGGCTGACGTGCTGTGCTCCAG 918
QY 919 CTGGATCGACG 929
    ||| ||| |||
Db 919 TTGACCCGCG 929
```

```
RESULT 6
US-09-823-373-15
; Sequence 15, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Acidovorax delafieldii
US-09-823-373-15
```

```
Query Match 27.3%; Score 283.8; DB 3; Length 1776;
Best Local Similarity 57.0%; Pred. No. 3e-54;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

QY 19 AAGTATCGCGCGCGCGGTGACGGCCGCGGTCTCTGATCTGACCGCACAGTC 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 350 AAGTTCCTCGCGCAACGTTTCAGGACAGACCGGTATGCTCGACGACGACGACATC 409
QY 79 GAGAAAGCATCGGCTGATCGAGCAGCGCGCAAGACAGACGTGCGCTGATCGCATTC 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 410 GACAAGTCGATCGGCATCATCGAAGAAGCTGCCAAAGGGCGCGAGTCTGATCGCTTTC 469
QY 139 CCAGAGACTTGATTCGGGCTATCCCTTTTGATATGGCTGGCGCGCGCTTGGGGC 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 470 CCGAAGTATTCAATTCGGGCTACCCCTATTGGGCGTGGCTCGGCGACGTGAAGTACAGC 529
QY 199 ATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGCTGCGCGGACGACGAGTGGCAG 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 530 CTAAGCTTACTTCACTATCAAGAAATTCGTTGAGCTAGGTAGCAGACCTATGCGT 589
QY 259 GCCCTGGCGATGCGCGCGCGCGCGCGCGCGCATGTCGTGCGCGGCTATAGCAGCGC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 590 CGCTTCAGCTGCGCGCGCGCGCAACAAATCGCACTGTCATGAGGCTATTTCGAGCGG 649
QY 319 GCGGCGCGCAGCTCTATATGGGCGCAGCGCATTCGCGCCCGATGCGCATGATCGCC 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 GAAGCCGATCGCGCTATCTGAGCCAGGTGTTCAATCGACGAGCGTGGCGAGATCGTTGCC 709
QY 379 GCGGCGCGCAGCTCAAGCTTACCCATGCGGAGCGCACCGTGTTCGGCGAGAGACGGC 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 710 AATCGCGCGCAAGCTGAAGCCCAACACACGTTGAGCGTACGATCTACGGCGAAGCAACGGA 769
```

```
QY 439 AGCATCTCGCGGTGCACGATACCCGCATCGGGCGCCTCGCGCGCTGTGCTGGAG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 770 ACCGATTTCTCAACGACGACTTCCGCTTCGGAACGCGTGGTGATTAAGTCTGCGGAA 829
QY 499 CACATCCAGCATTTGTGGAATACCCATGTACGCGCGCGACGAACAGGTCCAGTCCG 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 CATTTCAACCGCTCAGCAAGTTTATGATGTACAGCCTCGGTGAGCAGGTCCAGTGTGA 889
QY 559 TCGTGCCGAGCTTCAACCTCTATCGCGCATGCGCTATGCGTCCGACCGAGTCAAT 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 TCGTGCCGCGATGTCCCTCTTCAAGCCGATGTTTCCAACTGAGCATCGAAGCCAAC 949
QY 619 ACCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTACGTGCTGGCGTCCGCG 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 950 GCGACGGTCAACCCGCTCGTACGCAATCGAAGCCCAACCTTGTGTCTGCTCGACGAG 1009
QY 679 ACCGTTTCGCGGAGATGATCAAGGTATTGTGATACGCCCGACAGAGATGTTCTC 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1010 GTGATCGGACTTAGCGCGATCGAAACGTTCTGCTCAACGACGACGCGCACTGTTG 1069
QY 739 AAGCGCGCGCGGTTTGGCATGATTTTCGGCGCGCGCGCGCGCGCTGGCGGAGCCG 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1070 CCGAAGGATGTGCTGGCGCGCATTTACGGCCGATGGAAGCGAGCTTGCGAAGCCT 1129
QY 799 CTCCCGAGACCGAAGAGGAGTCTGCTGTCGCCGATATCGACCTCGGCGATGCGGTTG 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1130 CTGGCGGAAGATGTGAGGGGATCTTGATCGACAGAGATCGATCGACGAGATCTGCTG 1189
QY 859 GCCAAGCGCGCGCGCATCCGGCGGCACTATTTCACGCGCGCAAGCTAACCGCGCTGCTG 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1190 GCGAAGGCTGAGCCGATCCGGTCCGGCACTATTCCGGCGCTGACGTGCTGCTCCAG 1249
QY 919 CTGGATCGACG 929
    ||| ||| |||
Db 1250 TTGACCCGCG 1260
```

```
RESULT 7
US-09-823-373-16
; Sequence 16, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: A synthetic
; OTHER INFORMATION: version of the nitrilase gene
US-09-823-373-16
```

```
Query Match 19.4%; Score 202; DB 3; Length 1110;
Best Local Similarity 51.3%; Pred. No. 5.1e-36;
Matches 469; Conservative 0; Mismatches 445; Indels 0; Gaps 0;

QY 12 CATGACGAATATCGCGCGCGGTGACGGCGCGCGGTGTTCTCTGATCTCGACCG 71
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 12 CAACTCCAGTTCCTTGCGTCTACTGTTCAGCTGAGCCAGTTTGTGACGACGACGC 71
```



```
OY 72 CACAGTCGAGAAAGCGATCGCCCTGATCGAGCAGCGGCCCAAGCAGAGACGTGGCCTGAT 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 TACTATTGACAAGTCTATCGGTATCATCGAAGAGCTGCCAAAAGGGTGCTCTTGAT 131
OY 132 CGCATTCACAGAGACTTGATTCGCCGCTATCCCTTTGGATATGCTGGCGCGCCGCC 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 CGCTTCCAGAGATTTCATTCAGGTTACCACTAGGCGCTGTGGTGACGTTAA 191
OY 192 TTGGGGCATGCGCTTCGTCCAGCGCTATTTGAGAAATTCGCTCGTCGCGCGAGCAAGCA 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 GTACTCTTGTCTTACTTCTCCAGATATCACGAGAACTCTTTGAGTTGGTGACGACAG 251
OY 252 GTGGCAGGCTTGCGGAGATGCGCCCGCCGACGCGCATGCTGCTGGCGCGCTATAG 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 AATGCGTAGACTGCAATTGGCTGCCCGTAGAACAATAATGCTTTGTCATGGGTTATTC 311
OY 312 CGAGCGCGCGCGGAGCCTCTATATGGGCGCAGCGATCTTCGGCCCCGATGGCGATCT 371
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 CGAGAGAGAGCTGGATCTCGTTACTTGTCCCAAGTCTTCATCGACGAGAGAGTGAGAT 371
OY 372 GATCGCGCGCGCGGAGCTCAAGCCTACCCATGGCGAGCGCACCGTGTTCGGCGAGGG 431
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 TGTTCAAATCGTCGTAAGTTGAAGCCACTCAGTTGAGCGTACCATCTACGAGAGAAG 431
OY 432 AGACGCGACCACTCTCGCGGTGCACGATACCGCATCGGGCGCTCGCGCGCTCTGTTG 491
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 TAACGGAACCGATTCTTGACTACGACTTCGCTTCGGAAGAGTTGGTGATTTGAATG 491
OY 492 CTGGAGACATCCAGCCATTGTGCAATATACGCCATGTACCGCCGCGACGAACAGGTCCA 551
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 TTGGGAACATTTCCAACTCTGTCTAAGTTCATGATGTAATCTCTTGGTGAGCAAGTCCA 551
OY 552 CGTCGCGTCGTGCGCGAGCTTCAAGCTCTATCGCGCGCATGGCCTATGCGCTCGGAACGGA 611
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 CGTTGCTTCTTGGCCAGCTATGTCCCTCTTCAGCCAGATGTTTCCAAATGTCCATCGA 611
OY 612 GGTCAATACCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGCTACGTGCTGCGCTC 671
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 AGCCAACGCCACCGTCAACAGATCCTAAGCATGAAAGGTCAAACTTTGTCTTGTGCTC 671
OY 672 GTGCGCGACCGTTTCGCGGAGATGATCAAGTATGTGATACGCCGCAAGAGAGAT 731
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 TACCCAGGTCAATGGAACCTTCTGTATCGAAACCTTCTGTGAACGACGAACAGAGAGC 731
OY 732 GTTCCTCAAGCGCGCGGCTTTTCCATGATTTTCGGGCCGACGCGCGCGCTGCG 791
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 732 TTTGTTGCCAAGATGTGTTGGCAAGAAATTTACGGTCCAGATGATCTGAGCTTGC 791
OY 792 CGAGCCGCTCCCGGAGACCGAAGAGGAGTGTGTCGCCGATATCGACCTCGGCATGAT 851
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 CAAGCCTTTGGCTGAAGATGTGAGGGTATTTGTACGCTGAGATCGAATTGGAGCAAT 851
OY 852 CGCGTTGGCCAGCGCGCGCGGATCCGCGGGCCACTATTCACGGCCGACGTAACGCG 911
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 852 TCTGCTGGCCAGGCTGGAGCCGATCCAGTCGGTCACTACTCCAGACCTGACGCTTGTTC 911
OY 912 GCTGCTGCTGATC 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 912 CGTCCAGTTCGACC 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 8

```
US-09-823-373-3
; Sequence 3, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauban, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
```

```
; TITLE OF INVENTION: Acidovorax Facilis 72W
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Acidovorax facilis
US-09-823-373-3

Query Match 12.7%; Score 132.2; DB 3; Length 385;
Best Local Similarity 59.0%; Pred. No. 1.7e-20;
Matches 227; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

OY 165 CTTTGGATATGCTGGCGCGCGCTTGCGCATGCGCTTCCAGCGCTATTGCA 224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 CTATTGGCGGTGCTCGCGACGTGAAGTACAGCCTTAAGCTTTACTTCAAGCTATCAGCA 60
OY 225 GAATTCGTCGTGCGCGCAGCAGCAGTGGCAGCGCCCTGGCGATGCGCGCCGCCCA 284
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GAATTCGTTGAGACTAGTGACGACCGTATGCGTCGCTCCAGCTGGCGCGCGCCGCAA 120
OY 285 CGCATGCAATGTCGTGCGCGCTATAGCAGCGCGCGCGCGCGCGCTTATATGGCCA 344
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 CAAATCGCACTCGTCATGGGCTATTCCGAGCGGGAAGCCGATCGCGCTATCTGAGCCA 180
OY 345 GCGCATCTTCGCGCGCGCGCGCGCGCTGATCGCGCGCGCGCGCGCGCTCAAGCTACCCA 404
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GGTGTCATCGACGAGCGGTGGCGAGATCGTTGCCAATCGCGCGCAAGCTGAAGCCACACA 240
OY 405 TCGGAGCGCACCGTGTTCGGCGGAGAGACGGCAGCCATCTCGCGGTGCACGATACCGC 464
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 CGTTGAGCGTACGATCTACGCGCGAAGCAACGAAACCGATTCTCAACGACGACTTCGC 300
OY 465 CATCGGCGCGCTCGCGCGCTCTGTGCTGGGAGACATCCAGCCATTGTGAAATACGC 524
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 GTTCGAGCGCGTGGTGATTTGAAGTCTGGGAACATTTCCAAACGCTCAGCAAGTTTCAT 360
OY 525 CATGACGCGCGCGCAGCAACAGTTC 549
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 GATGACAGCCTCGGTGAGCAGTTC 385
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

## RESULT 9

```
US-08-447-702-4
; Sequence 4, Application US/08447702
; Patent No. 5629190
; GENERAL INFORMATION:
; APPLICANT: Petre, Dominique
; APPLICANT: Cerbeleaud, Edith
; APPLICANT: Levy-Schil, Sophie
; APPLICANT: Crouzet, Joel
; TITLE OF INVENTION: Polypeptides Possessing A Nitrilase
; TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,
; Patent No. 5629190
; TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them
; TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy-
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,702  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92-09-882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 87..1148  
US-08-447-702-4

Query Match 10.5%; Score 109.6; DB 2; Length 1194;  
Best Local Similarity 49.8%; Pred. No. 2.2e-15;  
Matches 277; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

QY 31 GCGGCGGTGACAGCGCGCGGCTGTTCTCGATCTGACCGCACAGTCGAGAAAGCGATC 90  
DB 114 GCAGCAGTGCAGCTGCTCCTGTATTGATGATGAGGCAACAGTAGATAAACTTGT 173  
QY 91 GGCCTGATCGACGAGCGCGCCAGCAGAGCTGCGCTGATCGCATTCGACAGACTTGG 150  
DB 174 AAGTTAATAGCAGACGACATCTATGGCGCCAGGTTATCGCTCCAGAAAGCATTT 233  
QY 151 ATTCGCGCTATCCCTTTTGATATGCTGGCGCGCGCTGGGCGATGCGCTTCTGTC 210  
DB 234 ATTCGCGCTATCCATATGATTTGACATCAATATGACTTCATGGAATGATGTGG 293  
QY 211 CAGCGCTATTCGAGAAATTCGCTCGTGGCGCGCAGCAGCAGTGGCGCCCTGGCGGAT 270  
DB 294 GCCGTCCTTTTCAAGAAATGCGATTGAATCCCAAGCAAGAAAGTTCAACAATTAGTAT 353  
QY 271 GCGGCGCGCGCCAGCGCATGATGCTGCGCGCGCTATAGCGAGCGCGCGCGCGCAGC 330  
DB 354 GCTGCAAAAAGAAATGAGTTTACGTTTTCGTTTCTGTATCAGAGAAAGATAATGCCCTCG 413  
QY 331 CTCTATATGGCCAGCGCATCTTCGCGCGCCGATGCGATCTGATCGCGCGCGCGCAG 390  
DB 414 CTATATTTGACGCAATTTGTGTTTGACCGCGAATGTAATTGATTGGCAAGCACAGAAA 473  
QY 391 CTCAAGCCTACCATGCGAGCGCACCGTTCGCGCGAGGAGAGACGCGCATCTCGCG 450  
DB 474 TTCAAGCCCACTAGTAGTAAGAGCTGTATGGGAGATGGGAGTGAAGCATGGCTCCC 533  
QY 451 GTGACGATACCGCCATCGGGCGCTCGCGCGCTGTGTGTTGGAGACATCCAGCCA 510  
DB 534 GTATTTAAACAGAGTATGGGAATCTTGGGGGACTCCAGTGCTGGGACATGCTCTCCA 593  
QY 511 TTGTGAAATACGCCATGTACGCGCGCGCAGAACAGGTCCACGTCGCGCTTGGCGGAGC 570  
DB 594 TTAACATTTGGCGCGATGGCTCATTTGAACGACAGGTACATGTTGCTTCTGCGCAGCC 653  
QY 571 TTCAGCCTTATCGCG 586

DB 654 TTCGTCCCTAAGCGG 669  
RESULT 10  
US-08-465-615-4  
Sequence 4, Application US/08465615  
Patent No. 5635391  
GENERAL INFORMATION:  
APPLICANT: PETRE, Dominique  
APPLICANT: CERBELEAUD, Edith  
APPLICANT: LEVY-SCHILL, Sophie  
APPLICANT: CROUZET, Joel  
TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,615  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
APPLICATION NUMBER: FR 9209882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 003025-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-615-4

QY 31 GCGGCGGTGACAGCGCGCGGCTTCTCGATCTGACCGCACAGTCGAGAAAGCGATC 90  
DB 114 GCAGCAGTGCAGAGCTGCTCCTGTATTGATGATGAGGCAACAGTAGATAAACTTGT 173  
QY 91 GGCCTGATCGACGAGCGCGCCAGCAGACGTCGCGCTGATCGCATTCGACAGACTTGG 150  
DB 174 AAGTTAATAGCAGAAAGCAGCATGTGGCGCCAGGTTATCGCTTCCAGAAAGCATTT 233  
QY 151 ATTCGCGCTATCCCTTTTGATATGCTGGCGCGCGCTTGGGCGATGCGCTTCTGTC 210  
DB 234 ATTCGCGCTATCCATATGATTTGACATCAATATGACTTCATGGAATGATGTGG 293  
QY 211 CAGCGCTATTCGAGAAATTCGCTCGTGGCGCGCAGCAGCAGTGGCAGCGCCCTGGCGAT 270

Db 294 GCCGCTCTTTTCAAGATGCGATTGAAATCCCAAGAAAGATTCAACAAATTAGTAT 353  
Qy 271 GCGCCCCCGCCAGCGCATGTCATGTCGTCGCGCTATAGCAGCGCGCGCGCAGC 330  
Db 354 GCTGCAAAAAGAAATGAGTTTACGTTTTCGTTTCTGTATCAGAGAAAAGTAATGCTCG 413  
Qy 331 CTCTATATGGCCAGCGCATCTTTCGCCCCCGATGCGCATGTCGCGCGCGCGCAAG 390  
Db 414 CTATATTTGACGCAATTGTGTTGACCCGAATGTAATTGATTGGCAGCAGAGAAA 473  
Qy 391 CTCAGCCTACCCATGCGGAGCGCACCGCTGTCGCGAGAGAGACGCGCATCTCGCG 450  
Db 474 TTAAAGCCCACTAGTAGTAAAGAGCTGTATGGGAGATGGGATGAAGCATGCTCCC 533  
Qy 451 GTGACGATACCGCCATGCGGCGCGCTGCTGTTGCTGGAGACATCCAGCCA 510  
Db 534 GTATTTAAACAGATATGGAATCTTGGGGGACTCCAGTGTGGAACATGCTCTCCCA 593  
Qy 511 TTGTGAAATACGCCATGTACGCGCGCGAGCAAGTCCACGTGCGTGGCGGAGC 570  
Db 594 TTAAACATGCGCGCATGCGCTCATTTGAACGAACAGGTACATGTTGCTTCTGCGCAGCC 653  
Qy 571 TTCAGCCTCTATCGC 586  
Db 654 TTCGTCCCTTAAAGCG 669

## RESULT 11

US-09-902-540-4902  
; Sequence 4902, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 4902  
; LENGTH: 1995  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-4902

Query Match 5.5%; Score 57; DB 3; Length 1995;  
Best Local Similarity 48.6%; Pred. No. 0.0013;  
Matches 258; Conservative 0; Mismatches 255; Indels 18; Gaps 3;

Qy 525 CATGTACGCGCCGACGACAGGTCCACGTCGCGTCGTGCGCGAGCTTCAAGCTTATCG 584  
Db 129 CGTGTGCGCGCGACGAGCGGTGGCAATTGCTCCTCCGCGAAGAGTCTTACCT 188  
Qy 585 CGGCATGGCTATGCGCTCGGACCGAGGTCAATACCGCGCAAGCAGATCTACGCGGT 644  
Db 189 CGTCATCGGGAAGATTCTGGAAGCGGCAAGACGTCGCGCGGAGGACCATCCACCGGG 248  
Qy 645 CGAGGGCGGCTGCTACGTGCTGCGCTGCGGACCGCTTTCGCGGAGATGATCAAGGT 704  
Db 249 CTACGGTTTCTTGTGCGAGAACGCGACTTTCGCGCGCTTTCGCGAGCGCGGCTGTGT 308  
Qy 705 ATTGGTGAATACGCGCGACAGAGATGTTCTCAAGCGCGCGCGGCTTTTGCAATGAT 764  
Db 309 GTTCATGCGCGCGAGCGCGAGCCATCACTGATGGCAACAGCGTCAGCGCAAGCT 368  
Qy 765 TTTTGGGCGC--CGACGGCGCGCGCTG-----GCCGAGCGCTCCCGAGACCGA 812  
Db 369 GCGCATGATTGCGCGCGCGCGCTGCGCTGCAATCTCTGCTACGAGCGGTCCGACCTGATGA 428

Qy 813 AGAGGACTGCTGTGTCGCGATATGCACTCGCATGATCGCTTGGCCAAAGCGCGGC 872  
Db 429 TGAGGCGCTGCGCGGTGAGAGGCGAGCGCATCGCTTCCGCTGATGTTCAAGCGCGGC 488  
Qy 873 CGATCCGGCGGCGCACTATTTCAGCGCC-----CGACGTAACGCGGCTGCTGTGATCG 926  
Db 489 GGGTGGCGGCGGCGCGCGCATGCGGCTGTGTGACACGAGCATCGAGCTTCGCGCGGCTT 548  
Qy 927 ACGTCCGCGCCCAACGCGTGTGACGCTTGATCGCGCATTCGAACCGCAAAACGAGACAA 986  
Db 549 GCGCGCGCGCGCTCGGAGCGCAGATGCTTTCGAGAGCGGAGGCTCATCTGAGAA 608  
Qy 987 GGGCAGCGCGCGCGCTGCGCGTGTGGCGGAAGCGCGCGCGCGCA 1037  
Db 609 GGCCTGATCGACGCGCGGACGTCGAGGTCCAGTCTTCCGCGACACGCA 659

## RESULT 12

US-09-902-540-1241/C  
; Sequence 1241, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1241  
; LENGTH: 28493  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(28493)  
; OTHER INFORMATION: unsure at all n locations  
US-09-902-540-1241

Query Match 5.5%; Score 57; DB 3; Length 28493;  
Best Local Similarity 48.6%; Pred. No. 0.002;  
Matches 258; Conservative 0; Mismatches 255; Indels 18; Gaps 3;

Qy 525 CATGTACGCGCGCGACGACAGGTCCACGTCGCGTCGTGCGCGAGCTTCAAGCTTATCG 584  
Db 2496 CGTGTGCGCGCGCGACGAGCGGTGGCAATTGCTCCTCCGCGAAGAGTCTTACCT 2437  
Qy 585 CGGCATGGCTATGCGCTCGGACCGGAGGTCAATACCGCGCAAGCAGATCTACGCGGT 644  
Db 2436 CGTCATCGGGAAGATTCTGGAAGCGGCAAGACGTCGCGCGGAGGACCATCCACCGGG 2377  
Qy 645 CGAGGGCGGCTGCTACGTGCTGCGCTGCGGACCGCTTTCGCGGAGATGATCAAGGT 704  
Db 2376 CTACGGTTTCTTGTGCGAGAACGCGGACTTTCGCGCGCTTTCGCGAGCGCGGCTGTGT 2317  
Qy 705 ATTGGTGAATACGCGCGACAGAGATGTTCTCAAGCGCGCGGCTTTTGCAATGAT 764  
Db 2316 GTTCATCGGCGCGAGCGCGAGCCATCACTGATGGCAACAGCGTCAGCGCAAGCT 2257  
Qy 765 TTTTGGGCGC--CGACGGCGCGCGCTG-----GCCGAGCGGCTCCCGAGACCGA 812  
Db 2256 GCGCATGATTGCGCGGCGGTGCGCTGCAATCTGCTGCTACGAGCGCTCCGACCTGATGA 2197  
Qy 813 AGAGGACTGCTGTGCGCGATATCGACCTCGCATGATCGCTTGGCCAAAGCGCGGC 872  
Db 2196 TGAGGCGCTGCGGTGAGGCGAGCGCATCGGCTTCCGCTGATGTTCAAGCGCGGC 2137  
Qy 873 CGATCCGCGCGCGCACTATTTCAGCGCC-----CGACGTAACGCGGCTGCTGTGATCG 926





```

Db      79360 GGCAGACACGGGATCGAGCCGCCATCCGGGCGGTCCGCCCGCTGCGCCGACGTACCGCT 79419
QY      732 GTTCC---TCAAGGCCGGCGGCGGTTTGGCCATGATTTTCGGGCCGACGGCCGCCCT 788
Db      79420 GGGCCCGGGGAGGTCTGTCGCGACGACGTGTCGTGTCTGACGTGACCCGGCCG 79479
QY      789 GAGCGAGCCGCTCCCGGAGACCGAAGAGGACTGCTGTGTCGCCGATATCGAAGCT--CGG 845
Db      79480 CGCCGTGTGACACCCGGCGGGGCCCCCGAGCCTGAAGGTGCGCTACACGCCCGCTGCACGG 79539
QY      846 CATGATCGCGTTGGCCCAAGGCGGCGCCGATCCGGCGGGCCACTATTCACGCGCCGACGT 905
Db      79540 CGTGGGCGGCGCGCTGTGACCGCGCTTCGCCCGCGCGCTTCGGCATCCCGCGCT 79599
QY      906 AACCGGCTGCTCTGTGATCGACGTCCCGCCCAACGCGTCTGTCACGCTTGATGCGCATT 965
Db      79600 GGTGCCGAGAGGCGGCTGCCGACCCGGACTTCGGACCGTACGCTTCCCAACCCGGA 79659
QY      966 CGAACCGCAAAACGAGACAAAGGCGGCGCGCGCTGCGCGGTGGCGGAAAGCGC 1025
Db      79660 GGAGCCGGGGCGGTGACCTCTGTCGCGCTCGCCGAGCGCACCGGGGCGGACCTGGC 79719
QY      1026 CGCCGCC 1032
Db      79720 GATCGCC 79726

```

```

RESULT 15
US-10-152-886-94
; Sequence 94, Application US/10152886
; Patent No. 6912470
; GENERAL INFORMATION:
; APPLICANT: ECOPIA BIOSCIENCES INC.
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDIYNE RING
; FILE REFERENCE: 3011-3US
; CURRENT APPLICATION NUMBER: US/10/152,886
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 94
; LENGTH: 5877
; TYPE: DNA
; ORGANISM: Streptomyces citricolor
US-10-152-886-94

```

```

Query Match      5.2%; Score 54.6; DB 3; Length 5877;
Best Local Similarity 42.6%; Pred. No. 0.0053;
Matches 343; Conservative 0; Mismatches 459; Indels 3; Gaps 1;

```

```

QY      229 TCCTCTGTCGGCGGCAAGCAAGTGGAGGCCCTGGCGGATGCGGCCCGCGCCACGGC 288
Db      4963 TGGTTCGGCGCTTCTCTGCCGACGACCAAGCTGTGGCCGACCCCGGCAACCCGGACGGC 5022
QY      289 ATGCATGTCGTGCGCGGCTATAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
Db      5023 ATGATGACAGCCATTCAGTGTGCTGCTGCGGACGCCACGCTGTGCTGCGGACAGCATCGAG 5082
QY      349 ATCTTCGGGCCCCGATGGCGATCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408
Db      5083 CGGCTTGCGCTGGCCGACCGGGCGGACCAAGACTCCGAGTACGTGTTCTCGACGCCCGG 5142
QY      409 GAGCGCACCGTGTTCGGCGAGGAGAGCGGACCGCATCTCGCGGTGACGATACCGCCATC 468
Db      5143 GAGCGCTCGACGAGCGGGGACACCTAAGTCTACGACCTCGACGTCCG--ACCCCTCG 5199
QY      469 GGGCGCTCGGCGCGCTCTGTGTCGTGAGACACATCCAGCCATTGCGAAATACGCCATG 528
Db      5200 GGAACGCTGTGAAACGCTGGAGGGGCTGGCCCTGTGTGCGCTCCGCAAAACGCGCGGG 5259

```

```

QY      529 TACGCCCGGACGAACAGTCCACGTCCGCTGCTGAGCCGAGCTTCAGCCTCTATCGCGC 588
Db      5260 GCCGGCCCGTGGGTGCCCCCGCATGTCTGGTTCTGTACCTGAGAGCGCGCCTGGAACGGTA 5319
QY      589 ATGGCTATGCGCTCGGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAG 648
Db      5320 CTGGGCGGAGCGCGCGGTGTCTGAAACCGCGCCCGGACCGCCGACCGCGGACCGAG 5379
QY      649 GCGGCTGCTACGTGTGCGGTCTGCGCGGACCGCTTTCGCCGAGATGATCAAGTATTG 708
Db      5380 GACCGCGCTCCCGTACCGAGACCGCGCTCGGCGGGCGCTGGCGCGCGCGGTGAAGCTG 5439
QY      709 GTGATACGCCCCGACCAAGAGATGTTCTCAAGGCGCGCGCGGCTTTTGCCATGATTTTC 768
Db      5440 CGCCACCGGCGGACGAGCGGACCGGAACTCGACGGCGGGCGCGGCTTGAGGACCGGACG 5499
QY      769 GGGCCGACGGCGCGCGCTGCGCGAGCGCGCTCCCGGAGACCGAAGGAGACTGTGTC 828
Db      5500 GTGTGCGCTCGCACGACGCGCGGACTGACCTCGCGGTGTGGCGCGGACGCGTCCG 5559
QY      829 GCCATATGACCTCGGCATGATCGCTTGCCCAAGCGCGCGCGGATCCGCGGCGCAC 888
Db      5560 TGCATGTGAGTGTGTCGCGGAGCGGACCGCGGAGACTGGAACGGGCTGTGGCGCC 5619
QY      889 TATTACGAGCCCGACGTAAAGCGGCTGCTGTGATCGACGTCCGCCCAACGCGTGTG 948
Db      5620 GGTGCGCTCGCCCTGCGGAACCTGTGGCAACCGAGCGGGCGAGGACCGGCGGTCCG 5679
QY      949 ACGCTGATCGCGCATTCGAACCGCAAAACGAGGACAAAGGCGGACCGCGCGCTGCGC 1008
Db      5680 GGCACCAAGGTGTGAGCGCGCTGGAAGTCTGCGCAAGGCGGTGCGACACGAGCGC 5739
QY      1009 GTGTGCGGAAAGCGCGCGCGCG 1033
Db      5740 CTGACGCTGAGCGCGCTCCACCCGG 5764

```

Search completed: April 27, 2006, 01:43:04  
Job time : 216.785 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:12:44 ; Search time 1001.99 Seconds  
(without alignments)  
8591.290 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1041  
Sequence: 1 atgtcgagcccatgacgaa.....gcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1041	100.0	1041	3 US-09-751-299-1	Sequence 1, Appl
2	1041	100.0	1041	6 US-10-146-772-383	Sequence 383, App
3	1041	100.0	1041	6 US-10-241-742-383	Sequence 383, App
4	1041	100.0	1041	7 US-10-440-523-383	Sequence 383, App
5	1041	100.0	1041	7 US-10-440-503-383	Sequence 383, App
6	1041	100.0	1041	7 US-10-461-925-383	Sequence 383, App
7	682	65.5	1026	6 US-10-146-772-33	Sequence 33, Appl
8	682	65.5	1026	6 US-10-241-742-33	Sequence 33, Appl
9	682	65.5	1026	7 US-10-440-523-33	Sequence 33, Appl
10	682	65.5	1026	7 US-10-440-503-33	Sequence 33, Appl
11	682	65.5	1026	7 US-10-461-925-33	Sequence 33, Appl
12	486	46.7	1062	6 US-10-146-772-211	Sequence 211, App
13	486	46.7	1062	6 US-10-241-742-211	Sequence 211, App
14	486	46.7	1062	7 US-10-440-523-211	Sequence 211, App
15	486	46.7	1062	7 US-10-440-503-211	Sequence 211, App
16	486	46.7	1062	7 US-10-461-925-211	Sequence 211, App
17	478.2	45.9	1056	6 US-10-146-772-341	Sequence 341, App
18	478.2	45.9	1056	6 US-10-241-742-341	Sequence 341, App
19	478.2	45.9	1056	7 US-10-440-523-341	Sequence 341, App
20	478.2	45.9	1056	7 US-10-440-503-341	Sequence 341, App
21	478.2	45.9	1056	7 US-10-461-925-341	Sequence 341, App
22	470	45.1	1050	6 US-10-146-772-377	Sequence 377, App
23	470	45.1	1050	6 US-10-241-742-377	Sequence 377, App

24	470	45.1	1050	7 US-10-440-523-377	Sequence 377, App
25	470	45.1	1050	7 US-10-440-503-377	Sequence 377, App
26	470	45.1	1050	7 US-10-461-925-377	Sequence 377, App
27	457.2	43.9	1065	6 US-10-146-772-101	Sequence 101, App
28	457.2	43.9	1065	6 US-10-241-742-101	Sequence 101, App
29	457.2	43.9	1065	7 US-10-440-523-101	Sequence 101, App
30	457.2	43.9	1065	7 US-10-440-503-101	Sequence 101, App
31	457.2	43.9	1065	7 US-10-461-925-101	Sequence 101, App
32	455.2	43.7	1077	6 US-10-146-772-169	Sequence 169, App
33	455.2	43.7	1077	6 US-10-241-742-169	Sequence 169, App
34	455.2	43.7	1077	7 US-10-440-523-169	Sequence 169, App
35	455.2	43.7	1077	7 US-10-440-503-169	Sequence 169, App
36	455.2	43.7	1077	7 US-10-461-925-169	Sequence 169, App
37	454.6	43.7	1017	6 US-10-146-772-197	Sequence 197, App
38	454.6	43.7	1017	6 US-10-241-742-197	Sequence 197, App
39	454.6	43.7	1017	7 US-10-440-523-197	Sequence 197, App
40	454.6	43.7	1017	7 US-10-440-503-197	Sequence 197, App
41	454.6	43.7	1017	7 US-10-461-925-197	Sequence 197, App
42	453.6	43.6	1017	6 US-10-146-772-249	Sequence 249, App
43	453.6	43.6	1017	6 US-10-241-742-249	Sequence 249, App
44	453.6	43.6	1017	7 US-10-440-523-249	Sequence 249, App
45	453.6	43.6	1017	7 US-10-440-503-249	Sequence 249, App

ALIGNMENTS

RESULT 1  
US-09-751-299-1  
; Sequence 1, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; APPLICANT: Chaplin, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; CURRENT FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
; OTHER INFORMATION: environmental sample  
; NAME/KEY: CDS  
; LOCATION: (1)..(1041)  
US-09-751-299-1

Query Match 100.0%; Score 1041; DB 3; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGTGAGGCCATGAGGAAGTATCGCGCGCGGTGACAGGCGCGGTTCCTC	60
DB	1	ATGTGAGGCCATGAGGAAGTATCGCGCGCGGTGACAGGCGCGGTTCCTC	60
QY	61	GATCTGACCGACAGTGCAGAAAGCATCGGCTGATCGAGCAGGCGCAAGCAGAC	120
DB	61	GATCTGACCGACAGTGCAGAAAGCATCGGCTGATCGAGCAGGCGCAAGCAGAC	120
QY	121	GTGCGCTGATCGATTCCAGAGACTTGATTCGGCTATCCCTTTTGATATGCTG	180
DB	121	GTGCGCTGATCGATTCCAGAGACTTGATTCGGCTATCCCTTTTGATATGCTG	180

```
QY 181 GCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTCGAAATTGCTGTCGC 240
Db 181 GCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTCGAAATTGCTGTCGC 240
QY 241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATGGCGCCCGCCAGCGCATGTCGTG 300
Db 241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATGGCGCCCGCCAGCGCATGTCGTG 300
QY 301 GCCGGCTATAGCAGCGCGCGGGCGGCGAGCCTTATATGGCCAGCGCATCTTCGGCCCC 360
Db 301 GCCGGCTATAGCAGCGCGCGGGCGGCGAGCCTTATATGGCCAGCGCATCTTCGGCCCC 360
QY 361 GATGGCGATCTGATCGCGCGCGCGCCGCAAGCTCAAGCTTACCCTATGCGAGCGCACCGTG 420
Db 361 GATGGCGATCTGATCGCGCGCGCGCCGCAAGCTTACCCTATGCGAGCGCACCGTG 420
QY 421 TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGC 480
Db 421 TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGC 480
QY 481 GCGCTCTGTTGCTGGGAGACACATCCAGCCATTGTGAAATACGCCATGACGCCCGCGAC 540
Db 481 GCGCTCTGTTGCTGGGAGACACATCCAGCCATTGTGAAATACGCCATGACGCCCGCGAC 540
QY 541 GAACAGGTCCACGTGCGCTGTCGTCGCGGAGTTCAAGCTTATCGCGGATGGCTATGCG 600
Db 541 GAACAGGTCCACGTGCGCTGTCGTCGCGGAGTTCAAGCTTATCGCGGATGGCTATGCG 600
QY 601 CTCGACCGGAGGTCAATACCGCCGAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
Db 601 CTCGACCGGAGGTCAATACCGCCGAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
QY 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCCCGGAGATGATCAAGGTATTGTGATACGCC 720
Db 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCCCGGAGATGATCAAGGTATTGTGATACGCC 720
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGTTTGGCATGATTTCCGGGCCCGACGGC 780
Db 721 GACAAGAGATGTTCTCAAGCGCGCGCGGTTTGGCATGATTTCCGGGCCCGACGGC 780
QY 781 CGCGCCCTGGCGCAGCCGCTCCCGGAGACCGAAGAGGAGTCTGTGTCGCCATATGCAC 840
Db 781 CGCGCCCTGGCGCAGCCGCTCCCGGAGACCGAAGAGGAGTCTGTGTCGCCATATGCAC 840
QY 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGGCGGATCCGCGGGCCACTATTACAGGCC 900
Db 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGGCGGATCCGCGGGCCACTATTACAGGCC 900
QY 901 GACGTACCGCGGCTGCTGTGATCGACGTCCGCGCCCAACGCGTCTCAAGCTTGATGCC 960
Db 901 GACGTACCGCGGCTGCTGTGATCGACGTCCGCGCCCAACGCGTCTCAAGCTTGATGCC 960
QY 961 GCATTTCGAACCGCAAAACGAGGACAAAGGCGCAGCGCGCGCTGCGGTGGCGGAA 1020
Db 961 GCATTTCGAACCGCAAAACGAGGACAAAGGCGCAGCGCGCGCTGCGGTGGCGGAA 1020
QY 1021 AGCGCGCGCGCGCGCGCAGTAG 1041
Db 1021 AGCGCGCGCGCGCGCGCAGTAG 1041
```

## RESULT 2

US-10-146-772-383  
; Sequence 383, Application US/10146772  
; Publication No. US20030124698A1

; GENERAL INFORMATION:

; APPLICANT: Short, Jay

; APPLICANT: Weiner, David

; APPLICANT: Chaplin, Jennifer

; APPLICANT: Chi, Ellen

; APPLICANT: Milan, Aileen

; APPLICANT: Desantis, Grace

; APPLICANT: Madden, Mark

```
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 383
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-383
```

Query Match 100.0%; Score 1041; DB 6; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGCTGTCAGCGCGCGGCTGTTCTC 60
Db 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGCTGTCAGCGCGCGGCTGTTCTC 60
QY 61 GATTCGACCGCACAGTGCAGAAAGCATCGGCTGATCGAGCAGCGCGCCAAAGCAGAC 120
Db 61 GATTCGACCGCACAGTGCAGAAAGCATCGGCTGATCGAGCAGCGCGCCAAAGCAGAC 120
QY 121 GTGCGCTGATCGCATTCGAGAGACTTGATTCGCGCTATCCCTTTGGATATGCTG 180
Db 121 GTGCGCTGATCGCATTCGAGAGACTTGATTCGCGCTATCCCTTTGGATATGCTG 180
QY 181 GCGCGCGGCTTGGGCGATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGTGCGC 240
Db 181 GCGCGCGGCTTGGGCGATGCGCTTCGTCAGCGCTATTTCGAGAATTGCTGTGCGC 240
QY 241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATGCGCGCGCGCCAGCGCATGTCGTG 300
Db 241 GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATGCGCGCGCGCCAGCGCATGTCGTG 300
QY 301 GCCGGCTATAGCAGCGCGCGCGCGCAGCCTTATATGGCCAGCGCATCTTCGGCCCC 360
Db 301 GCCGGCTATAGCAGCGCGCGCGCGCAGCCTTATATGGCCAGCGCATCTTCGGCCCC 360
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCTATGCGAGCGCACCGTG 420
Db 361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCTATGCGAGCGCACCGTG 420
QY 421 TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACAGATACCGCCATCGGGCGCTCGGC 480
Db 421 TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACAGATACCGCCATCGGGCGCTCGGC 480
QY 481 GCGCTCTGTTGCTGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540
Db 481 GCGCTCTGTTGCTGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCCCGCGAC 540
QY 541 GAACAGGTCCACGTGCGCTGTCGTCGCGGAGTTCAAGCTTATCGCGGATGGCTATGCG 600
Db 541 GAACAGGTCCACGTGCGCTGTCGTCGCGGAGTTCAAGCTTATCGCGGATGGCTATGCG 600
QY 601 CTCGACCGGAGGTCAATACCGCCGAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
Db 601 CTCGACCGGAGGTCAATACCGCCGAGCCAGATCTACGCGGTGAGGGCGGCTGTAC 660
```

Db 601 CTCGGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACCGCGTCGAGGGCGCTGTAC 660  
Qy 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCCGAGATGATCAAGGATATTGGATACGCC 720  
Db 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCCGAGATGATCAAGGATATTGGATACGCC 720  
Qy 721 GACAAGAGATGTTCTCAAGCGCGCGGTTTGGCATGATTTTCGGGCCGAGCGC 780  
Db 721 GACAAGAGATGTTCTCAAGCGCGCGGTTTGGCATGATTTTCGGGCCGAGCGC 780  
Qy 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCCGAGAAGGAGTGTGTGCGCGCATATCGAC 840  
Db 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCCGAGAAGGAGTGTGTGCGCGCATATCGAC 840  
Qy 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGCGGCACATATTCACGGCCC 900  
Db 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGCGGCACATATTCACGGCCC 900  
Qy 901 GACGTAACGCGGCTGCTGTGATCGACGTCGCCGCCAAACGCGTCTGACGCTTGATGCC 960  
Db 901 GACGTAACGCGGCTGCTGTGATCGACGTCGCCGCCAAACGCGTCTGACGCTTGATGCC 960  
Qy 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGCGCTGCGCGTGGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGCGCTGCGCGTGGCGGAA 1020  
Qy 1021 AGCGCCGCCGCCGCGCAGTAG 1041  
Db 1021 AGCGCCGCCGCCGCGCAGTAG 1041

RESULT 3  
US-10-241-742-383

; Sequence 383, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 383  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-383

Query Match 100.0%; Score 1041; DB 6; Length 1041;

Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGTCCGAGCCCATGACGAAGTATCCGCGCGCGCGGTGCAAGGCCGCCGCTTCTC 60  
Db 1 ATGTCCGAGCCCATGACGAAGTATCCGCGCGCGCGGTGCAAGGCCGCCGCTTCTC 60  
Qy 61 GATCTGACCCGACAGTCCGAGAAAGCGATCCGCTGATCGAGCAGCGGCCCAAGCAGGAC 120  
Db 61 GATCTGACCCGACAGTCCGAGAAAGCGATCCGCTGATCGAGCAGCGGCCCAAGCAGGAC 120  
Qy 121 GTGCGCTGATCGCATTTCCAGAGACTTGGATTCCCGGCTATCCCTTTTGGATATGCTG 180  
Db 121 GTGCGCTGATCGCATTTCCAGAGACTTGGATTCCCGGCTATCCCTTTTGGATATGCTG 180  
Qy 181 GCGCGCGCGCTTGGGCGATGCGCTGCTCCAGCGCTATTTCGAGAAATTCGCTGCGC 240  
Db 181 GCGCGCGCGCTTGGGCGATGCGCTGCTCCAGCGCTATTTCGAGAAATTCGCTGCGC 240  
Qy 241 GGCAGCAAGCAGTGGCAGAGCCCTGGCGGATGCGGCCGCCCGCCACGCGCATGCTG 300  
Db 241 GGCAGCAAGCAGTGGCAGAGCCCTGGCGGATGCGGCCGCCCGCCACGCGCATGCTG 300  
Qy 301 GCCGCTATAGCAGAGCGCGCGCGCGCCTCTATATGGGCCAGCGATCTTCGCCCC 360  
Db 301 GCCGCTATAGCAGAGCGCGCGCGCGCCTCTATATGGGCCAGCGATCTTCGCCCC 360  
Qy 361 GATGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCAATCGGAGCGCACCGTG 420  
Db 361 GATGCGATCTGATCGCGCGCGCGCGCAAGCTCAAGCCTACCAATCGGAGCGCACCGTG 420  
Qy 421 TTCGCGAGGAGAGACGGCAGCCATCTCGCGGTGACAGATACCGGCCTCGGC 480  
Db 421 TTCGCGAGGAGAGACGGCAGCCATCTCGCGGTGACAGATACCGGCCTCGGC 480  
Qy 481 GCGCTCTGTGCTGGAGACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCAG 540  
Db 481 GCGCTCTGTGCTGGAGACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCAG 540  
Qy 541 GAACAGGTTCACGTCGCGTGTGCGGAGCTTACGCTCTATCGCGCATGGCCTATGCG 600  
Db 541 GAACAGGTTCACGTCGCGTGTGCGGAGCTTACGCTCTATCGCGCATGGCCTATGCG 600  
Qy 601 CTCGACCGGAGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGCTGTAC 660  
Db 601 CTCGACCGGAGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGCTGTAC 660  
Qy 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCCGAGATGATCAAGGATATTGGATACGCC 720  
Db 661 GTGCTGGCGTCTGCGCGCAGCCGTTTCGCCGAGATGATCAAGGATATTGGATACGCC 720  
Qy 721 GACAAGAGATGTTCTCAAGCGCGCGGTTTGGCATGATTTTCGGGCCGAGCGC 780  
Db 721 GACAAGAGATGTTCTCAAGCGCGCGGTTTGGCATGATTTTCGGGCCGAGCGC 780  
Qy 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCCGAGAAGGAGTGTGTGCGCGCATATCGAC 840  
Db 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCCGAGAAGGAGTGTGTGCGCGCATATCGAC 840  
Qy 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGCGGCACATATTCACGGCCC 900  
Db 841 CTCGGCATGATCGCGTTGGCCAAAGCGCGCGCGCATCCGGCGGCACATATTCACGGCCC 900  
Qy 901 GACGTAACGCGGCTGCTGTGATCGACGTCGCCGCCAAACGCGTCTGACGCTTGATGCC 960  
Db 901 GACGTAACGCGGCTGCTGTGATCGACGTCGCCGCCAAACGCGTCTGACGCTTGATGCC 960  
Qy 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGGCCGCCGCTGCGCGTGGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAACGAGACAAGGGCGAGCGCGGCCGCCGCTGCGCGTGGCGGAA 1020  
Qy 1021 AGCGCCGCCGCCGCGCAGTAG 1041  
Db 1021 AGCGCCGCCGCCGCGCAGTAG 1041



Db 1021 AGCGCCGCCGCCGCGAGTAG 1041

RESULT 4  
US-10-440-523-383

```
; Sequence 383, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitriases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 383
; LENGTH: 1041
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-383
```

Query Match 100.0%; Score 1041; DB 7; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTCGAGGCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTTCCTC 60
    |||||||
Db 1 ATGTCGAGGCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTTCCTC 60

QY 61 GATCTCGACCGACAGTGCAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGGAC 120
    |||||||
Db 61 GATCTCGACCGACAGTGCAGAAAGCGATCGGCTGATCGAGCAGCGGCCAAGCAGGAC 120

QY 121 GTGCGCTGATCGCATTTCCAGAGACTTGATTTCCCGCTATCCCTTTTGATATGCTG 180
    |||||||
Db 121 GTGCGCTGATCGCATTTCCAGAGACTTGATTTCCCGCTATCCCTTTTGATATGCTG 180

QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCCAGACGCTATTTGAGAAATTCGCTCGTGGC 240
    |||||||
Db 181 GCGCGCGCGCTTGGGGCATGCGCTTCCAGACGCTATTTGAGAAATTCGCTCGTGGC 240

QY 241 GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGGCGCGCGCGCGCAGCGCATGCTGCTG 300
    |||||||
Db 241 GGCAGCAAGCAGTGGCAGGCGCTGGCGGATGGCGCGCGCGCGCAGCGCATGCTGCTG 300

QY 301 GCGGCTATAGCAGGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCGCCCC 360
    |||||||
Db 301 GCGGCTATAGCAGGCGCGCGCGCGCGCGCTTATATGGGCCAGCGCATCTTCGCCCC 360

; QY 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCTACCCATGCGGAGCGCACCGTG 420
```

```
Db 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCTACCCATGCGGAGCGCACCGTG 420
    |||||||
QY 421 TTCGGCAGAGGAGACGGCAGCCATCTCGCGGTGACAGGATACCGCCATCGGGCGCTCGGC 480
    |||||||
Db 421 TTCGGCAGAGGAGACGGCAGCCATCTCGCGGTGACAGGATACCGCCATCGGGCGCTCGGC 480

QY 481 GCGCTCTGTTGCTGGAGACATCCAGCCATTGTGGAATAGCCATGTACGCCGCCGAC 540
    |||||||
Db 481 GCGCTCTGTTGCTGGAGACATCCAGCCATTGTGGAATAGCCATGTACGCCGCCGAC 540

QY 541 GAACAGGTCCACGTGCGCGTGGCGGAGCTTCAGCCCTTAATCGCGCATGGCCTATGCG 600
    |||||||
Db 541 GAACAGGTCCACGTGCGCGTGGCGGAGCTTCAGCCCTTAATCGCGCATGGCCTATGCG 600

QY 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACCGGTGAGGGCGGTGCTAC 660
    |||||||
Db 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTACCGGTGAGGGCGGTGCTAC 660

QY 661 GTGCTGCGCTGTCGCGGACCGTTTCGCCGAGATGATCAAGGTATTGGTATACGCC 720
    |||||||
Db 661 GTGCTGCGCTGTCGCGGACCGTTTCGCCGAGATGATCAAGGTATTGGTATACGCC 720

QY 721 GACAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGGCCGACGCG 780
    |||||||
Db 721 GACAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGGCCGACGCG 780

QY 781 CGGCGCTGGCCGAGCGGCTCCCGGAGACCCGAAGAGGAGACTGTGTCGCCATATGAC 840
    |||||||
Db 781 CGGCGCTGGCCGAGCGGCTCCCGGAGACCCGAAGAGGAGACTGTGTCGCCATATGAC 840

QY 841 CTCGGCATGATCGCGTTGGCCAGGCGCGCGCGCGATCCGGCGGGCCACTATTACAGGCC 900
    |||||||
Db 841 CTCGGCATGATCGCGTTGGCCAGGCGCGCGCGCGATCCGGCGGGCCACTATTACAGGCC 900

QY 901 GACGTAAAGCGGCTGCTGTCGATCGACGTCCCGGCCAACCGCTGTCACGCTTGATGCC 960
    |||||||
Db 901 GACGTAAAGCGGCTGCTGTCGATCGACGTCCCGGCCAACCGCTGTCACGCTTGATGCC 960

QY 961 GCATTGCAACCGCAAAACGAGACAAAGGGCGAGCGCGCGCGCTGCGGTGTCGGGAA 1020
    |||||||
Db 961 GCATTGCAACCGCAAAACGAGACAAAGGGCGAGCGCGCGCGCTGCGGTGTCGGGAA 1020

QY 1021 AGCGCCGCCGCCGCGCAGTAG 1041
    |||||||
Db 1021 AGCGCCGCCGCCGCGCAGTAG 1041
```

RESULT 5

```
US-10-440-503-383
; Sequence 383, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 383
; LENGTH: 1041
; TYPE: DNA
```

```

; ORGANISM: Unknown
;
; FEATURE:
;
; OTHER INFORMATION: Obtained from an environmental sample
;
US-10-440-503-383

```

Query Match	100.0%;	Score 1041;	DB 7;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.8e-276;		
Matches 1041; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	ATGTCGGAGCCCCATGACGAAGTATCGCGGCGCGGGGTGCAGAGCCGCGCGGTGTTCTC	60
Db	1	ATGTCGGAGCCCCATGACGAAGTATCGCGGCGCGGGGTGCAGAGCCGCGCGGTGTTCTC	60
QY	61	GATCTCGACCCGACAGTCGAGAAAGCATCGGCCTGATCGAGCAGGCGGCCAAGCAGGAC	120
Db	61	GATCTCGACCCGACAGTCGAGAAAGCATCGGCCTGATCGAGCAGGCGGCCAAGCAGGAC	120
QY	121	GTGCGCCTGATCGCATTCCACAGACTTGGATTCCCGGCTATCCCTTTGGATATGGCTG	180
Db	121	GTGCGCCTGATCGCATTCCACAGACTTGGATTCCCGGCTATCCCTTTGGATATGGCTG	180
QY	181	GGCGCGCCGGCTTGGGCGATGCGCTTCGTCCAGCGCTATTTCGAGATTGCTCGTCGC	240
Db	181	GGCGCGCCGGCTTGGGCGATGCGCTTCGTCCAGCGCTATTTCGAGATTGCTCGTCGC	240
QY	241	GGCAGCAAGCAGTGGCAGGCGCTGCGGATCGGCGCCCGCGCCACGCGCATGCAATGTCGTG	300
Db	241	GGCAGCAAGCAGTGGCAGGCGCTTGGCGGATCGGCGCCCGCGCCACGCGCATGCAATGTCGTG	300
QY	301	GCCGGCTATAGCGAGCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGAGCGCACCGTG	360
Db	301	GCCGGCTATAGCGAGCGCGCGCGCGCGCAAGCTTATATGGCCAGGCGCATCTTCGAGCCC	360
QY	361	GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGAGCGCACCGTG	420
Db	361	GATGGCGATCTGATCGCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGAGCGCACCGTG	420
QY	421	TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGGC	480
Db	421	TTCGGCGAGGAGACGCGCAGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGGC	480
QY	481	GCGCTCTGTTGCTGGAGCACATCCAGCCATTGTCCGAATACGCCATGTACGCCCGCAC	540
Db	481	GCGCTCTGTTGCTGGAGCACATCCAGCCATTGTCCGAATACGCCATGTACGCCCGCAC	540
QY	541	GAAACAGSTCCACGTCGCGTCTGGCGCCAGGCTTCAGCCTCTATCGCGGCATGGCTATGCG	600
Db	541	GAAACAGSTCCACGTCGCGTCTGGCGCCAGGCTTCAGCCTCTATCGCGGCATGGCTATGCG	600
QY	601	CTCGGACCCGAGGTCATATACCGCCGACAGCCAGATCTACGCGTGAAGGCGGCTGTAC	660
Db	601	CTCGGACCCGAGGTCATATACCGCCGACAGCCAGATCTACGCGTGAAGGCGGCTGTAC	660
QY	661	GTGCTGGCGTCGTGCGGACCGTTTCGCGCGAGATGATCAAGTATTGGTGATACGCC	720
Db	661	GTGCTGGCGTCGTGCGGACCGTTTCGCGCGAGATGATCAAGTATTGGTGATACGCC	720
QY	721	GACAAGGAGATGTTCTCTCAAGGCCGCGCGCGGTTTGCCATGATTTTCGGGCCCGACGCC	780
Db	721	GACAAGGAGATGTTCTCTCAAGGCCGCGCGCGGTTTGCCATGATTTTCGGGCCCGACGCC	780
QY	781	CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGGACTGCTGTCGCCGATATCGAC	840
Db	781	CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGGACTGCTGTCGCCGATATCGAC	840
QY	841	CTCGGCATGATCGCGTTGGCCAAAGGCGGCGCGGATCCGGCGGCGCCACTATTTCAGGGCC	900
Db	841	CTCGGCATGATCGCGTTGGCCAAAGGCGGCGCGGATCCGGCGGCGCCACTATTTCAGGGCC	900
QY	901	GACGTAACGCGGCTGCTGTGATGACGTCGCGGCCCAACGCGTCTCACGCTTGATGCC	960
Db	901	GACGTAACGCGGCTGCTGTGATGACGTCGCGGCCCAACGCGTCTCACGCTTGATGCC	960

QY	961	GCATTGGAACCGGCAAAACGAGGACAAAGGGCCAGCCGCCGCTGCGCGTGTGGCGGAA	1020
Db	961	GCATTGGAACCGGCAAAACGAGGACAAAGGGCCAGCCGCCGCTGCGCGTGTGGCGGAA	1020
QY	1021	AGCGCCGCGCGCGCGCAGTAG	1041
Db	1021	AGCGCCGCGCGCGCGCAGTAG	1041

RESULT 6  
US-10-461-925-383

```

: Sequence 383, Application US/10461925
: Publication No. US20040053378A1
: GENERAL INFORMATION:
: APPLICANT: Mark J. Burk
: APPLICANT: Desantis, Grace
: APPLICANT: Morgan, Brian
: APPLICANT: Zhu, Zoulin
: TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
: FILE REFERENCE: 09010-270001
: CURRENT APPLICATION NUMBER: US/10/461,925
: PRIOR FILING DATE: 2003-06-13
: PRIOR APPLICATION NUMBER: US 60/389,317
: PRIOR FILING DATE: 2002-06-13
: PRIOR APPLICATION NUMBER: US 60/392,944
: PRIOR FILING DATE: 2002-06-28
: NUMBER OF SEQ ID NOS: 386
: SOFTWARE: FastSeq for Windows Version 4.0

```

Query Match	100.0%;	Score 1041;	DB 7;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.8e-276;		
Matches 1041;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

OY		1	ATGTCGAGGCCCATGACGAATGATCGCGCGCGCGGTGCAGGCCGGCGGTCTC	60
Db	1	ATGTCGAGCCCATGACGAATGATCGCGCGCGGTGCAGGCCGGCGGTCTC	60	
OY		61	GATCTCGACCGCACAGTCCGAAGAAGCATCGGCTGATCGAGCAGCGGCCAAGCAGGAC	120
Db	61	GATCTCGACCGCACAGTCCGAAGAAGCATCGGCTGATCGAGCAGCGGCCAAGCAGGAC	120	
OY		121	GTGCGCCTGATCGCATTTCCCAAGACTTGATTCCCCTATCCCTTTGGATATGGCTG	180
Db	121	GTGCGCCTGATCGCATTTCCCAAGACTTGATTCCCCTATCCCTTTGGATATGGCTG	180	
OY		181	GCGCGCGCGCTTGGGCGATGCGCTTCGTCAGCGCTATTTGCAATTGCTGTCGC	240
Db	181	GCGCGCGCGCTTGGGCGATGCGCTTCGTCAGCGCTATTTGCAATTGCTGTCGC	240	
OY		241	GCGACGAAGCAGTGGCAGGCCCTGGCGGATGCGGCCGCCAGCGCATGTCGTG	300
Db	241	GCGACGAAGCAGTGGCAGGCCCTGGCGGATGCGGCCGCCAGCGCATGTCGTG	300	
OY		301	GCCGCTATAGCGAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCCATCTTCGCCCC	360
Db	301	GCCGCTATAGCGAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGGCCATCTTCGCCCC	360	
OY		361	GATGGCGATCTGATCGCCCGCGCGCGCAAGCTCAAGCCTAACCTATGCGAGCGCACCGTG	420
Db	361	GATGGCGATCTGATCGCCCGCGCGCGCAAGCTCAAGCCTAACCTATGCGAGCGCACCGTG	420	
OY		421	TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGC	480
Db	421	TTCGGCGAGGAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGGGCGCCTCGGC	480	
OY		481	GCGCTGTGTTGCTGGAGCATCCAGCCATTGTCGAATAACGCATGTACGCCGCCGAC	540

```

Db      481 GCGCTGTGCTGGAGACATCCAGCCATTGCGAATACGCCATGTAAGCCGCGAC 540
QY      541 GAACAGGTCCACGTGCGCTGCTGGCCGAGCTTCAAGCTTATCGCGGATGCTATGCG 600
Db      541 GAACAGGTCCACGTGCGCTGCTGGCCGAGCTTCAAGCTTATCGCGGATGCTATGCG 600
QY      601 CTGAGCCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660
Db      601 CTGAGCCGAGGTCAATACCGCCGCAAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660
QY      661 GTGCTGCGCTGCTGCGCGACCGCTTTCGCGAGATGATCAAGGTATGATGATGATG 720
Db      661 GTGCTGCGCTGCTGCGCGACCGCTTTCGCGAGATGATCAAGGTATGATGATGATG 720
QY      721 GACAAAGAGATGTTCTCAAGGCGCGCGGCTTTCGCGATGATTTTCGCGCGCGACGCG 780
Db      721 GACAAAGAGATGTTCTCAAGGCGCGCGGCTTTCGCGATGATTTTCGCGCGCGACGCG 780
QY      781 CGCGCCCTGGCGGAGCGCGCTCCGAGAGACCGAAGGAGACTGTGTCGCGCATATCGAC 840
Db      781 CGCGCCCTGGCGGAGCGCGCTCCGAGAGACCGAAGGAGACTGTGTCGCGCATATCGAC 840
QY      841 CTGCGCATGATCGCGTGGCCCAAGCGCGCGCGATCCGCGGCGCACTATTCAGCGGCC 900
Db      841 CTGCGCATGATCGCGTGGCCCAAGCGCGCGCGATCCGCGGCGCACTATTCAGCGGCC 900
QY      901 GACGTAACGCGGCTGCTGCTGATCGACGCTCCGCGCCCAACGCGCTGCTGATGCC 960
Db      901 GACGTAACGCGGCTGCTGCTGATCGACGCTCCGCGCCCAACGCGCTGCTGATGCC 960
QY      961 GCATTGCAACCGCAAAACGAGGACCAAGGCGACGCGCGCTGCGGTGCGGAA 1020
Db      961 GCATTGCAACCGCAAAACGAGGACCAAGGCGACGCGCGCTGCGGTGCGGAA 1020
QY      1021 AGCGCGCGCGCGCGCAGTAG 1041
Db      1021 AGCGCGCGCGCGCGCAGTAG 1041

```

```

RESULT 7
US-10-146-772-33
; Sequence 33, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1026

```

```

; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-33

Query Match      65.5%; Score 682; DB 6; Length 1026;
Best Local Similarity 79.4%; Pred. No. 2.5e-177;
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY      1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGGCGGTGACGCGCGCGGTTCCTC 60
Db      1 ATGTTAAGTCCGTGACGACGATGCGCGCGCGCGGTGACGCGCGCGGTTCCTC 60
QY      61 GATCTGACCGCACAGTCCGAGAAAGCATCGCGCTGATCGACGAGCGCGCAAGCAGAC 120
Db      61 GATCTGACCGCACAGTCCGAGAAAGCATCGCGCTGATCGACGAGCGCGCAAGCAGAT 120
QY      121 GTGCGCTGATCGCATTCCTCAAGACTTGATTCCTCGGCTATCCCTTTGATATGCTG 180
Db      121 GTGCGCTGATCGCGTTTCCGAAACCTGATTCCTCGGCTATCCGCTGATCTGCTC 180
QY      181 GCGCGCGCGCTTGGGCGATGCGCTTCTGTCAGCGCTATTCGAAATTCGTCGTGCGC 240
Db      181 GCGCGCGCGCTTGGGCGATGCGCTTCTGTCAGCGCTATTCGAAATTCGTCGTGCGC 240
QY      241 GCGAGCAAGCAGTGGCAGCGCTGCGGATGCGCGCGCGCGCGCATGATGCTGTC 300
Db      241 GCGAGCAACAGTGAACGCGATCGCCGATGCGCGCGCGCGCGCATGACCGCTGTC 300
QY      301 GCGGCTATAGCGAGCGCGCGCGCGCAGCTCTATATGCGCAGCGCATCTTGGCCCC 360
Db      301 GTCGCTTCAAGCAGCGCGCGCGCGCAGCGCTCTACATGCGCAGCGCATCTTGGCCCC 360
QY      361 GATGCGCATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCCCATGCGAGCGCACCGT 420
Db      361 GAGGCGAGCTCATCGCGCGCGCGCGCAAGCTCAAGCGCGCACACCGCGAGCGAGT 420
QY      421 TTCGCGAGGAGGAGACGCGCAGCGCTGCGGTGACGATACCGCATCGGCGCTCGGC 480
Db      421 TTCGCGAGGAGGAGACGCGCAGCGCTGCGGTGACGATACCGCATCGGCGCTCGGC 480
QY      481 GCGCTCTGTTGCTGGGAGCACATCCAGCCATTGTGAAATACGCCATGTACGCGCGCAC 540
Db      481 GCGCTCTGTTGCTGGGAGCACATCCAGCCCTCTCGAAATACCGCATGTATGCGCGCAC 540
QY      541 GAACAGTCCACGTCGCTGCGCGCGCGCGAGCTTACGCTTATCGCGCATGCGCTATGCG 600
Db      541 GAACAGTCCACGTCGCTGCGCGCGCGCGAGCTTACGCTTATCGCGCATGCGCTATGCG 600
QY      601 CTGAGCCGAGGTCAATACCGCGCGCGAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660
Db      601 CTGAGCCGAGGTCAATACCGCGCGCGAGCCAGATCTACGCGGTGAGGCGGCTGTAC 660
QY      661 GTGCTGCGCTGCTGCGCGACCGCTTTCGCGGAGATGATCAAGGTATGATGATGATG 720
Db      661 GTGCTGCGCTGCTGCTGCGCGACCGCTTTCGCGGAGATCTGATGATGATGATGATG 720
QY      721 GACAAAGAGATGTTCTCAAGGCGCGCGCGGCTTTCGCGATGATTTTCGCGCGCGACGCG 780
Db      721 GACAAAGAGATGTTCTCAAGGCGCGCGCGGCTTTCGCGATGATTTTCGCGCGCGACGCG 780
QY      781 CGCGCCCTGGCGGAGCGCGCTCCGAGAGACCGAAGGAGACTGTGTCGCGCATATCGAC 840
Db      781 CGCGCCCTGGCGGAGCGCGCTCCGAGAGACCGAAGGAGACTGTGTCGCGCATATCGAC 840
QY      841 CTGCGCATGATCGCGTGGCCCAAGCGCGCGCGCGCTGATCGACGAGCGCGCGCGCGCC 900
Db      841 CTGCGCATGATCGCGTGGCCCAAGCGCGCGCGCGCTGATCGACGAGCGCGCGCGCGCC 900
QY      901 GACGTAACGCGGCTGCTGCTGATCGACGCTCCGCGCGCAACGCGCTGATCGCTGATGCC 960
Db      901 GACGTAACGCGGCTGCTGCTGATCGACGCTCCGCGCGCGCGCTGATCGCTGATGCC 960

```







; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-33

Query Match 65.5%; Score 682; DB 7; Length 1026;
Best Local Similarity 79.4%; Pred. No. 2.5e-177;
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60
DB 1 ATGTTAAGTCCCGTGAACGAGTATCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60
QY 61 GATCTGACCGCACAGTCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAC 120
DB 61 GATCTGACCGCACCGCTCGAGAAAGCGATCGATCATCGAGCAGCGCGCAAGCAGAT 120
QY 121 GTGCGCTGATCGATTCCTCCAGAGACTTGATTCCTCCGCTATCCCTTTGATATGCTG 180
DB 121 GTGCGCTGATCGCTTTCCTCGAAACCTGATTCCTCCGCTATCCCTCTGATCTGCTC 180
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGC 240
DB 181 GCGTCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGC 240
QY 241 GGCAGCAGCAGTGGGAGGCGCCCTGGCGGATGCGCGCGCGCGCGCGCATGTCGTG 300
DB 241 GGCAGCAACAGTGGAAACGCGATCGCGCGCGCGCGCGCGCGCATGACCGTCTC 300
QY 301 GCGGCTATAGGAGCGCGCGCGCGCGCGCTTATATGGGCGAGCGCATCTTGGCCCC 360
DB 301 GTCGGCTTACGAGCGCGCGCGCGCGCGCTTACATGGGCGAGCGCATCTTGGCCCC 360
QY 361 GATGCGCATGATCGCGCGCGCGCGCAAGCTCAAGCCTACCCATGCGGAGCGCACCGTG 420
DB 361 GAAAGCGAGCTCATCGCGCGCGCGCGCAAGCTCAAGCGACACACCGCGAGCGAAGTG 420
QY 421 TTGCGCGAGGAGACGCGCGCATCTCGCGGTGACAGTACCGCGCATCGCGCGCTCGC 480
DB 421 TTGCGCGAGGCGAGCGCGCATCTTGGCGTTTACGAGACGGCGGTGTCGATCGCG 480
QY 481 GCGCTCTGTGCTGGGAGACATCCAGCGCTTGCMAATACGCGATGTACGCGCGCAC 540
DB 481 GCGCTCTGTGCTGGGAGACATCCAGCGCTTGCMAATACGCGATGTATGCGCGCAAC 540
QY 541 GAACAGGTCCAGTGCCTGCTGCGCGCGAGCTTCAAGCTTATCGCGCATGCGCTATGCG 600
DB 541 GAACAGGTGATGTGCGCTGCTGCGCGCGAGCTTCAAGCTTATCGCGCATGCGCTATGCG 600
QY 601 CTGCGACGCGAGTCAATACCGCGCGAGCAGATCTACGCGGTGAGGCGCGCTGTAC 660
DB 601 CTGCGCGGAGGTGAACACCGCGCGAGCAGTCTACGCGGTGAGGCGCGCTGTAC 660
QY 661 GTGCTGCGGTGTCGCGCGACCGTTTCGCGGAGATGATCAAGGTATGTTGATACGCGCC 720
DB 661 GTGCTGCGCTCTGTCTCTGTCTGTGACACCGAGATCTGAAGGTGCTGATGACACGCGCC 720
QY 721 GACAAAGAGATGTTCTCAAGCGCGCGCGGTTTGGCCATGATTTTGGGCGCGACGCGC 780
DB 721 GACAAAGAGCGTTGCTGCTGCGCGCGGCGGTTTCTGATGATCTTGGCGCGACGCGC 780
QY 781 CGCGCGCTGCGCGAGCGCTCCGCGAGACCGAAGAGGAGTGTGCTGCGCGATATGAC 840

DB 781 CGCGCGCTCGCCCAAGCGCGCTGCCGAGACCGAAGAGGGCTGTCTACGCGCGAGATCGAT 840
QY 841 CTCGCGATGATCGCGTTGGGCCAAGCGCGCGCGCGATCCGCGCGCGCACTATTACAGCGCC 900
DB 841 CTCGCGCGATCGCGCTCGCCCAAGCGCGCGCGCGATCCGCGCGCGCACTTACGCGCGCC 900
QY 901 GACGTAACGCGCTGCTGTGATCGACGTCCGCGCGCAAGCGCGGTGTCAGCGTTGATGCC 960
DB 901 GACGTGACGCGGTGTTGCTGAACCGCGCGCGCGCGCGCGCGCGGTGTCAGCGTTGATGCC 960
QY 961 GCATTGCAACCGCAAAACGAGACAGGCGCGACCGCGCGCGCGCGCGCGCGGTGTCGCG 1018
DB 961 CGCTTCGAGGTCTGTGACAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018

RESULT 10

US-10-440-503-33
; Sequence 33, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chl, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-33

Query Match 65.5%; Score 682; DB 7; Length 1026;
Best Local Similarity 79.4%; Pred. No. 2.5e-177;
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTC 60
DB 1 ATGTTAAGTCCCGTGAACGAGTATCGCGCGCGGTGACAGCGCGGTGTTCTC 60
QY 61 GATCTGACCGCACAGTGAAGACGATCGGCTGATCGAGCAGCGCGCAAGCAGAC 120
DB 61 GATCTGACCGCACCGCTCGAGAAAGCGATCGCGATCATCGAGCAGCGCGCAAGCAGAT 120
QY 121 GTGCGCTGATCGATTCCTCCAGAGACTTGATTCCTCCGCTATCCCTTTGATATGCTG 180
DB 121 GTGCGCTGATCGCTTTCCTCGAAACCTGATTCCTCCGCTATCCCTCTGATCTGCTC 180
QY 181 GCGCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGC 240
DB 181 GCGTCGCGCGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGC 240
QY 241 GGCAGCAGCAGTGGGAGGCGCCCTGGCGGATGCGCGCGCGCGCGCGCATGTCGTG 300
DB 241 GGCAGCAACAGTGGAGACGCGATCGCGCGCGCGCGCGCGCGCATGACCGTCTGTC 300
QY 301 GCGGCTATAGCAGCGCGCGCGCGCGCGCTTATATGGGCGAGCGATCTTGGCGCC 360
DB 301 GTCGGCTTACGAGCGCGCGCGGAGGCGCGCTTACATGGGCGAGCGATCTTGGCGCC 360



US-10-146-772-211

```
; Sequence 211, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 211
; LENGTH: 1062
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-211
```

Query Match 46.7%; Score 486; DB 6; Length 1062;  
 Best Local Similarity 67.8%; Pred. No. 1.7e-123;  
 Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

```
QY 17 CGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGTTCTCGATCTCGACCGACAG 76
    |||||
Db 17 CGAATATACAAGTGGCCGCGGTGACAGCGCGCGGTGTTCTCGACCTGACCGCTCGG 76
    |||||
QY 77 TCGAAGAAAGCGATCGGCTGATCGAGAGCGCGCGCAAGCAGAGAGCGCGCTGATCGCAT 136
    |||||
Db 77 TCGAAAAGCGCTCGGTTTCATCGACGAGCGCGCGCGCGCGCGCGCTCATCGCCT 136
    |||||
QY 137 TCCAGAGACTTGATTTCCCGGCTATCCCTTTTGATATGCTGGCGCGCGCTTGGG 196
    |||||
Db 137 TTCCGAGACTGGATACCCGGTACCCCTGGTGGATCTGGCTAGGCGCGCGCGCTGGG 196
    |||||
QY 197 GCAT-----GCGCTTCTCCAGCGCTATTTTGAGATTCGCTCGCGCGCGCAGCAGC 250
    |||||
Db 197 CTATCATGCGCGCTTCTGCTCGCGCTATTTTGACACTCGCTCAGCTACGACAGCCCGC 256
    |||||
QY 251 AGTGCAGGCGCTTGGCGGATGCGCGCGCGCGCGCGCGCGCATGCGATCGTGCGCGCTATA 310
    |||||
Db 257 AGGCCGAGAGCTCGCGCGCGCGCGCGCGCGCGCAAGCAGAGATGCTGGTGGTCTCGCCTCT 316
    |||||
QY 311 GCGAGCGCGCGCGCGCGCGCGCTCTATATGGGCGAGCGCGATCTTCGCGCGCGCGATGCGATC 370
    |||||
Db 317 CCGAGCGCGAGCGCGCGCGCGCTTACATCGCGCAATGATCATCGCGCGCGAGCGGAGAA 376
    |||||
QY 371 TGATCGCGCGCGCGCGCGCGAGCTCAAGCTTACCCCATGCGGAGCGCGACCGTGTTGCGGAGG 430
    |||||
Db 377 CCATCGCCCAAGCGCGCGCGAGCTCAAGCCGACCCACGCGGAGCGCGACCGTGTTGCGGAGG 436
    |||||
QY 431 GAGACGCGAGCCATCTCGCGGTGACAGATACCGCCCATCGGCGCGCTCGGCGCTGTGTT 490
    |||||
Db 437 GCGACGGCTCGCATCTTGCGGTGACAGACTTGATGTGGCGCGCTCGGCGCGCTGTGCT 496
    |||||
```

```
QY 491 GCTGGAGACATCCAGCCATTGTCGAATACGCCATGTACGCCCGCGAGCAAGGTCC 550
    |||||
Db 497 GCTGGAGACACTGACAGCCGCTGTCCAATACGCCATGTATGCGAGAACGAAAGGTGC 556
    |||||
QY 551 ACGTCGGCTGCTGGCGCGAGCTTACGCTCTATCGCGCGCATGGCTTATCGGCTCGGACCGG 610
    |||||
Db 557 ATGTCGGGCTGCGCGAGCTTTTCCCTTACGATCCGTTGCGGACCGGCTCGGCGCGG 616
    |||||
QY 611 AGTCAATACCGCGCGAGCGAGCTATACGCGGTGAGGGCGGCTGCTACGTGTCGGCT 670
    |||||
Db 617 AAGTAAATATGCGCGCGAGCAAAATCTATGCGGTGAGGGCTGCTGTTTCATCGCGCC 676
    |||||
QY 671 CGTGGCGAGCCGTTTCGCGCGAGATGATCAAGTATTGTTGATATCGCCCGAGCAAGAGA 730
    |||||
Db 677 CGTGGCGAGCCGTTTCGAGGCGATGATCGAGAACTCTGCGATACCGCGAGAGCATC 736
    |||||
QY 731 TGTTCCTCAAGGCGCGCGGCTTTTGCCATGATTTTGGGCGCGAGCGCGCGCTGG 790
    |||||
Db 737 AGTTCCTGATGCGCGCGGCTTTTGCGGATTTACGCGCGCGCGCGCGCGCTCG 796
    |||||
QY 791 CCGAGCGCTCCCGGAGACCGGAGGGAAGGACTGCTGCTGCGGATATCGACCTCGGATGA 850
    |||||
Db 797 CCGCGCGCTGCGCGCGCGCGCGAGAGGAGGCTTGTCTACCGCGAGCATGATCTCGGATGA 856
    |||||
QY 851 TCGCGTTGGCCCAAGGCGCGCGCGATCCGCGCGCGCGCACTATTCACGCGCGAGTAAAGC 910
    |||||
Db 857 TTTGCGTTGCCAAAGCGCGCGAGCTGCGCGCGCGCGCATTAATGCAAGCGCGCGATCACCC 916
    |||||
QY 911 GCGTGTGCTGATCGACGTCGCGCGCGCGCGCGCGCGCTGCTGACGCTTGATCGCGATTGAAAC 970
    |||||
Db 917 GCGTGTGTTCAACAATCGCGCTGGGTATCGGGTGAACCATGCGCGTTGCGGATGATG 976
    |||||
QY 971 CGCAAAACGAGACCAAGGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCG 1030
    |||||
Db 977 CCGAGACCAAGCGCGAGACCGGCTTAAGCCGGAACCCCAAGGCAAGCGGCGCGCGT 1036
    |||||
QY 1031 CCGCGCAG 1038
    |||||
Db 1037 TCGCGCG 1044
    |||||
```

RESULT 13  
 US-10-241-742-211  
 ; Sequence 211, Application US/10241742  
 ; Publication No. US20040002147A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Short, Jay  
 ; APPLICANT: Weiner, David  
 ; APPLICANT: Chaplin, Jennifer  
 ; APPLICANT: Chi, Ellen  
 ; APPLICANT: Milan, Aileen  
 ; APPLICANT: Desantis, Grace  
 ; APPLICANT: Madden, Mark  
 ; APPLICANT: Burk, Mark  
 ; TITLE OF INVENTION: Nitrilases  
 ; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
 ; CURRENT APPLICATION NUMBER: US/10/241,742  
 ; CURRENT FILING DATE: 2002-09-09  
 ; PRIOR APPLICATION NUMBER: US/10/146,772  
 ; PRIOR FILING DATE: 2002-05-15  
 ; PRIOR APPLICATION NUMBER: US 60/309,006  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US 60/351,336  
 ; PRIOR FILING DATE: 2002-01-22  
 ; PRIOR APPLICATION NUMBER: US 60/300,189  
 ; PRIOR FILING DATE: 2001-06-21  
 ; PRIOR APPLICATION NUMBER: US 09/751,299  
 ; PRIOR FILING DATE: 2000-12-28  
 ; PRIOR APPLICATION NUMBER: US 60/254,414  
 ; PRIOR FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/173,609  
 ; PRIOR FILING DATE: 1999-12-29









Db 797 CGGCGCCGCTGCGCCCGCCGACCAAGGAAGGCTTGCTCTACGCCGACATCGATCTCGGATGA 856  
Qy 851 TCGCGTTGGCCCAAGGCGGCGCGCGGATCCGGCGGGCCACTATTTCACGGCCCGACGTACGC 910  
Db 857 TTTCGGTTGCCAAAGCGGCGAGCCGATCCGGCCGGGCATTATGACGCGCCCGACGTCACCC 916  
Qy 911 GGCTGCTGCTGATCGACGTCGCCGCCAACGCGTCGTACGCGCTTGATGCCGATTCGAAC 970  
Db 917 GGCTTCTGTCAACAATCGGCTGCGTATCGGGTCGAGACCATGGCGTTGCCGATCGATG 976  
Qy 971 CGCAAAACGAGACAAAGGCGGACGCGCCCGCGCTGCGCGTGTGGCGGAAGAAGCGCCCG 1030  
Db 977 CGGAGACCAAGGCGGAAGCACCGGCTTAAGCCGGAACCAAGGCACCGAACGTGGCGCGCT 1036  
Qy 1031 CCGCGCAG 1038  
Db 1037 TCGCGCCG 1044

Search completed: April 27, 2006, 00:01:33  
Job time : 1004.99 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:27:34 ; Search time 626.12 Seconds  
(without alignments)  
6752.353 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1041  
Sequence: 1 atgtcggagcccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA\_New:\*

1: /SIDS5/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
2: /SIDS5/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /SIDS5/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
4: /SIDS5/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
5: /SIDS5/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
6: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
7: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq1:\*  
8: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq2:\*  
9: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq3:\*  
10: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq4:\*  
11: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
12: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq2:\*  
13: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq3:\*  
14: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq4:\*  
15: /SIDS5/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	297.4	28.6	1071	9	US-10-537-075-6 Sequence 6, Appli
2	283.8	27.3	1110	9	US-10-919-182-3 Sequence 3, Appli
3	283.8	27.3	1110	9	US-10-919-182-7 Sequence 7, Appli
4	283.8	27.3	1110	9	US-10-919-182-11 Sequence 11, Appli
5	282.2	27.1	1110	9	US-10-919-182-5 Sequence 5, Appli
6	282.2	27.1	1110	9	US-10-919-182-13 Sequence 13, Appli
7	282.2	27.1	1110	9	US-10-919-182-15 Sequence 15, Appli
8	282.2	27.1	1110	9	US-10-919-182-17 Sequence 17, Appli
9	50.6	4.9	9975	7	US-10-496-351-4 Sequence 4, Appli
10	50.6	4.9	82746	7	US-10-496-351-56 Sequence 56, Appli
11	48.2	4.6	1706	11	US-11-096-568A-20685 Sequence 20685, A
12	47.8	4.6	113193	7	US-10-478-943D-1 Sequence 1, Appli
13	47.6	4.6	2247	7	US-10-496-351-8 Sequence 8, Appli
14	47.6	4.6	4770	7	US-10-496-351-3 Sequence 3, Appli
15	47.6	4.6	8192	9	US-10-933-746-36 Sequence 36, Appli
16	47	4.5	720	13	US-11-079-906-1 Sequence 1, Appli
17	47	4.5	720	14	US-11-079-476-1 Sequence 1, Appli
18	47	4.5	726	14	US-11-175-690-115 Sequence 115, App

19	47	4.5	1314	14	US-11-069-642-46	Sequence 46, Appl
20	47	4.5	1314	14	US-11-069-642-48	Sequence 48, Appl
21	47	4.5	1314	14	US-11-069-642-50	Sequence 50, Appl
22	47	4.5	1314	14	US-11-069-642-52	Sequence 52, Appl
23	47	4.5	1314	14	US-11-069-642-54	Sequence 54, Appl
24	47	4.5	1314	14	US-11-069-642-56	Sequence 56, Appl
25	47	4.5	1314	14	US-11-069-642-58	Sequence 58, Appl
26	47	4.5	1314	14	US-11-069-642-60	Sequence 60, Appl
27	47	4.5	1314	14	US-11-069-642-62	Sequence 62, Appl
28	47	4.5	1539	14	US-11-032-236-5	Sequence 5, Appli
29	47	4.5	1781	14	US-11-108-890A-11	Sequence 11, Appl
30	47	4.5	2349	11	US-11-194-991-76	Sequence 76, Appl
31	47	4.5	2355	11	US-11-194-991-78	Sequence 78, Appl
32	47	4.5	2502	9	US-10-537-971-1	Sequence 1, Appli
33	47	4.5	3660	14	US-11-032-236-3	Sequence 3, Appli
34	47	4.5	4151	14	US-11-213-368-15	Sequence 15, Appl
35	47	4.5	4862	14	US-11-082-154A-87	Sequence 87, Appl
36	47	4.5	4894	14	US-11-181-148-2	Sequence 2, Appli
37	47	4.5	5041	14	US-11-193-750-6	Sequence 6, Appli
38	47	4.5	5094	9	US-10-948-344-1	Sequence 1, Appli
39	47	4.5	5162	14	US-11-082-154A-26	Sequence 26, Appl
40	47	4.5	5510	14	US-11-082-154A-71	Sequence 71, Appl
41	47	4.5	6119	14	US-11-082-154A-126	Sequence 126, App
42	47	4.5	6748	8	US-10-655-872-3	Sequence 3, Appli
43	47	4.5	7350	8	US-10-655-872-8	Sequence 8, Appli
44	47	4.5	7487	14	US-11-181-148-4	Sequence 4, Appli
45	47	4.5	7600	14	US-11-082-154A-115	Sequence 115, App

ALIGNMENTS

RESULT 1  
US-10-537-075-6  
; Sequence 6, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesselier, Maria  
; APPLICANT: Zielinski, Thomas  
; APPLICANT: Hauser, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 1071  
; TYPE: DNA  
; ORGANISM: Alcaligenes faecalis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1068)  
; OTHER INFORMATION: coding for nitrilase  
US-10-537-075-6

Query Match 28.6%; Score 297.4; DB 9; Length 1071;  
Best Local Similarity 58.2%; Pred. No. 2.3e-65;  
Matches 523; Conservative 0; Mismatches 376; Indels 0; Gaps 0;

QY	25	CGCGCGCGCGCGGTGACGCGCGCGCGGTGTTCTTCGATCTGCACCGCACAGTCGAGAA	84
DB	22	CGGCGAGCCGCGCTACAGCGCGCTTCCCAACTACGATCTGGCAACGGGTGTGATPAA	81
QY	85	GCGATCGCCTGATGAGCAGCGCGCAAGCAGCAGCGCTGATCGATTCAGAG	144
DB	82	ACCATTTGAGCTGCTGTCAGCGCGCGGATGAGGCGTGTGACCTGATCGTTGGTGAA	141
QY	145	ACTTGATTCCCGCTATCCCTTTTGATATGCTGGCGCGCGGCTTGCGCATGCC	204



```

; LOCATION: (1)..(1110)
; US-10-919-182-3
142 ACCTGGCTGCCCCGGATATCCCTTCCACGCTCTGGCTGGCGCACCCGGCCTGGTCTGAA 201
QY 205 TTGCTCCAGCGCTATTTCGAGAAATTCGCTCGTGGCGGGCAGCAAGCAGTGGCAGCCCTG 264
Db 202 TACAGTGGCCGCTACTATGCAACTGCTCTCGCTGGACAGTGACAGATTTCACGCATT 261
QY 265 GCGGATGCGGCCCCCGCCAGCGGCATGCAATGCTGGCCGCTATAGCGAGCGCGGC 324
Db 262 GCGGAGCGCGCACGGAAGCTTGGGTATTTTCATGCACTGGGTATAGCGAGCGCAGCGC 321
QY 325 GCGAGCCTCTATATGAGCCAGCGCATCTTGGCCCCGATGGCGATCTGATCGCCGCGC 384
Db 322 GCGAGCCTTTACTTGGGCCAATGCTGATGACGACAGGCGAGATGCTGTGCTGCGT 381
QY 385 CGCAAGCTCAAGCTTACCCATGCGAGCGCAGCGTGTTCGGCGAGGAGACGCGCAT 444
Db 382 CGCAAACTCAAAACCCAGCATGTAGAGCGCAGCTATTGGTGAAGTTATGCCCGTAT 441
QY 445 CTGCGGTGACAGATACCGCCATCGGGCGCTGGCGCGCTCTGTGCTGGGAGACATC 504
Db 442 CTGATTTGTGTCGACACAGAACTGGGACGCGCTGCTCTATGCTGTGGGAGCATTTG 501
QY 505 CAGCCATTGTCAAAATACGCCATGTACGCCCGCCGACGACAGAGTCCAGTTCGCTGTG 564
Db 502 TCGCCCTTGAAGCAATACCGCGCTGTACTCCAGCATGAAGCCATTCACTTGCCTGG 561
QY 565 CCGAGCTTCAAGCTCTATTCGGCGGCATGGCCTTATGCGCTCGGACCGGAGTCAATACGCC 624
Db 562 CCGTCTTTTCGCTATACAGCGAAGAGGCCACGCCCTCAGTGCCAAAGTGAACATGCT 621
QY 625 GCAAGCCAGATCTACCGGCTCGAGGGCGGCTGCTACGCTGCGCTGCTGGCGCAGCCGT 684
Db 622 GCCTCGCAATCTATTCTGTTGAAGGCCAGTGTCTTACCATGCGCGCAGCAGTGTGTC 681
QY 685 TCGCCGAGATGATCAAGGTATTGTGATACGCCCGACAAAGAGATGTTCTCAAGGCC 744
Db 682 ACCCAAGAGAGCGCTAGACATGCTGGAAGTGGTGAACACAAAGCCCTTGTGAAGTG 741
QY 745 GCGGCGGTTTGGCATGATTTTCGGCGCCGACGCGCGCTGGCGAGCCGCTCCG 804
Db 742 GCGGCGGCAATTCCATGATTTTGGCGCGGACGAGCAGCACATGGCTCCCTACCTGCT 801
QY 805 GAGACCGAAGAGGAGCTGCTGTGCGCGATATCGACCTCGGCATGATCGCGTGGCCAG 864
Db 802 CACGATGCGGAGGGCTTGATCATTTGCCGATCTGAATATGAGAGATGCTTCGCCAA 861
QY 865 GCGGCGGCGCATCCGGCGGCGCATATCAAGCGCCGACGTAACGCGCTGCTGCTGA 923
Db 862 GCGATCAATGACCCCGTAGGCCACTATTTCAAAACCGAGGCCACCCGCTGCTGCTGA 920

```

```

RESULT 2
US-10-919-182-3
; Sequence 3, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: C12584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Acidovorax facilis 72W
; FEATURE:
; NAME/KEY: CDS

```

```

; LOCATION: (1)..(1110)
; US-10-919-182-3
Query Match 27.3%; Score 283.8; DB 9; Length 1110;
Best Local Similarity 57.0%; Pred. No. 6.1e-62;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;
QY 19 AAGTATGCGGCGCGCGGTGAGCGCCGCGGTGTTCTCGATCTCGACCGCAGATC 78
Db 19 AAGTCTCGCGGCAACCGTTGAGGAGAGCCGCTATGGCTCGACGCAAGCAGATC 78
QY 79 GAGAAAGCATCGGCTGATCGACAGCGCGCCAGCAGACGTCGCCCTGATCGCATTC 138
Db 79 GACAAGTCGATCGGCATCATCGAAGAAAGTCCCAAAAGGCGCGAGTCTGATCGCTTC 138
QY 139 CCAGAGCTTGGATTCCCGGCTATCCCTTTGGATATGGCTGGCGCGCGCTTGGGGC 198
Db 139 CCGAAGTATTATTCCTCGGCTAACCCCTATTGGCGCTGCTGGCGAGCTGAAGTACAGC 198
QY 199 ATGCGCTTCTCGACCGCTATTTCGAGAAATTCGCTCGCGCGCGCAGCAAGCAGTGCAG 258
Db 199 CTAACTTTTACTTACGCTATCAAGAAATTCGTTGAGCTAGGTGACGACCGTATGCTG 258
QY 259 GCCCTGCGGATGCGGCGCGCGCGCCAGCGCATGCAATGCTGCGCGCTATAGCGAGCC 318
Db 259 CGCTTCAAGCTGCGCGCGCGCGCAAAATTCGACTGCTCATGCGCTATTCGAGCGG 318
QY 319 GCGGCGGAGACCTCTATATGAGCGCAGCGCATCTTGGCCCCCGATCGCATTCGCTG 378
Db 319 GAAAGCGGATCGCGCTATGAGCCAGGTGTTTATGACGAGCGTGGCGAGATCGTTGCC 378
QY 379 GCGGCGGCAAGCTCAAGCTTACCCATGCGAGCGCACCGTTCGGCGAGGAGACGCC 438
Db 379 AATGCGGCAAGCTGAAGCCACACACGTTGAGCGTACGATCTACGCGGAGGCAACGA 438
QY 439 AGCATCTCGCGGTGACAGATACCGCCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAG 498
Db 439 ACCGATTTCTCACGACGACTTCGCGTTCGAGCGCGCTCGGATGATGAACTGCTGGAA 498
QY 499 CACATCCAGCCATTGTCAAAATACGCCATGTACCGCGCCGACGAAAGGTCCAGTCCG 558
Db 499 CATTTCCAACCGCTCAGCAAGTTTATGATGTATACGCTCGGTGAGCAGGTCCAGTTGCA 558
QY 559 TCGTGGCGGAGCTTCACTCTATTCGCGCATGCGCTATGCGCTCGGACCGGAGTCAAT 618
Db 559 TCGTGGCGGCGGATGTCCTCTTTCAGCCGAGTGTTCCTCAACTGAGCATGGAAGCCAA 618
QY 619 ACCGCGGCAAGCCAGATCTACCGGTCGAGGCGCGCTCTACGCTGCGCTGCTGCGG 678
Db 619 GCGAGGTCACCCGCTGCTGATCGCAATCGAAGGCCAAACCTTGTGCTTGTGCTGACG 678
QY 679 ACCGTTTCGCGGAGATGATCAAGGTATTGTGATACCGCCGACAAAGAGATGTTCTC 738
Db 679 GTGATCGAGCTAGCGGATCGAAACGTTCTGCTCAACGACGAAACAGCGGCACTGTTG 738
QY 739 AAGGCGGCGGCGGTTTTCATGATTTTCGGCGCGACGCGCGCGCTGGCGAGCCG 798
Db 739 CCGCAAGATGTGCTGGCGCGCATTTTACGGCGCGGATGGAAGCGAGCTTGGAAAGCCT 798
QY 799 CTCCCGGAGACCGAAGGAGACTGCTGTGCGCGATATCGACCTCGGCATGATCGCGTTG 858
Db 799 CTGCGGGAAGATGCTGAGGGGATCTTGTACGCGAGATCGATCTGAGCAGATTTGCTG 858
QY 859 GCCAAGCGCGCGCGATCCGCGCGCGCATATTCACGCGCGCGCATACCGCGCTGCTG 918
Db 859 GCGAAGGCTGAGCGCGATCCGCTGCGGCACTATTCGCGCGCTGACGCTGCTGCTCAG 918
QY 919 CTGATCGAGC 929
Db 919 TTGACCCGCG 929

```

RESULT 3

```
US-10-919-182-7
; Sequence 7, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B4
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-7
```

```
Query Match      27.3%; Score 283.8; DB 9; Length 1110;
Best Local Similarity 57.0%; Pred. No. 6.1e-62;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;
```

```
OY 19 AAGTATCGCGCGCGCGGTGACAGCGCGCGCGCGGTCTCGATCTGACCGCACAGTC 78
    |||||
DB 19 AAGTCTCTCGCGCAACCGTTACAGGACAGAGCCGGTATGGCTCGACGACGACGACATC 78
OY 79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGCGCAAGCAGACGTCGCTGATCGCATTC 138
    |||||
DB 79 GACAAAGTCGATCGCATCATCGAAGAGCTGCCAAAGCGCGCGAGTCTGATCGCTTTC 138
OY 139 CCAGAGACTTGATTCGCGCTATCCCTTTGATATGCTGGCGCGCGCGCTGGGGC 198
    |||||
DB 139 CCGAAGTATTCAATCCGGGCTACCCCTATTGGCGCTGGCTCGGCGACGTGAAGTCAGC 198
OY 199 ATGCGCTTCGTCACGCGCTATTTGAGAATTGCTCGTGGCGGACAGCAGTGGCAG 258
    |||||
DB 199 CTAAGCTTACTTCACGCTATCACGAGATTCGTTGAGACTAGTGACGACCGTATGCCGT 258
OY 259 GCCCTGGCGGATCGGCGCGCGCGCGCGCGCATGATGTCGTGGCGCGCTATAGCGAGCGC 318
    |||||
DB 259 CGCCTCCAGCTGGCGCGCGCGCGCGCAACAAATCGCACTGTCATGGGCTATTCGAGCGCG 318
OY 319 GCGGCGGCGAGCCTTATATGCGCGCGCGCGCGCATTTCCGCCCGCGCATCTGATCGCC 378
    |||||
DB 319 GAAGCCGATCGCGCTATCTGAGCCAGGTGTTATCGACGAGCGTGGCGAGATCGTGGCC 378
OY 379 GCGCGCGCGCAAGCTCAAGCCTTACCCATGCGGACCGCACCGTGTTCGCGGAGAGAGCGC 438
    |||||
DB 379 AATCGCGCAAGCTGAGCCCAACACAGCTTGAGCGTACGATCTACGCGGAGGCAACGGA 438
OY 439 AGCCATCTCGCGGTGACGATACCGCCATCGGCGCGCTCGCGCGCTCTGTGCTGGAG 498
    |||||
DB 439 ACCGATTCTCACGACGACTTCGCGTTCGACGCGCTCGGTGATGAACTGCTGGAA 498
OY 499 CACATCCAGCCATGTCGAATAACGCCATGTACGCCCGCGGACGACGAGTCCAGTCCG 558
    |||||
DB 499 CATTTCAACCGCTCAGCAAGATCATGATGTACAGCTCTGGTGAGCAGGTCCAAGTTGCA 558
OY 559 TCGTGCGCGAGCTTCAAGCTCTATCGCGGCGATGGCTATGCGCTCGGACCGGAGGTCAAT 618
    |||||
DB 559 TCGTGCGCGGCGATGTCCCTCTTCAGCCGAGTGTTCCTCAACTGAGCATGGAAGCCAAC 618
OY 619 ACCGCGGCAAGCCAGATCTACGCGGTGAGGCGCGGTGCTACGTCGTGGCGTGGCGCG 678
    |||||
DB 619 GCGAGCGTCAATCCGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGCTGACGCGAG 678
```

```
OY 679 ACCGTTCCGCGGAGATGATCAAGTATGTGTGATACGCCCCGACAGAGATGTTCTC 738
    |||||
DB 679 GTGATCGGACCTAGCGCGATCGAAACGTTCTGCTCAACGACGACAGCGCACTGTTG 738
OY 739 AAGCGCGCGCGGCTTTTGCCATGATTTCCGGGCGCGCGCGCGCGCTGGCCGAGCG 798
    |||||
DB 739 CCGCAAGGATGTGGCTGGCGCGCGCATTTACGCGCCGATGGAAGCAGCTTCGAAGCCT 798
OY 799 CTCCCGGAGACCGAAGAGGACTGCTGCTGCGCCGATATCGACCTTCGCATGATCGCGTTG 858
    |||||
DB 799 CTGGCGGAGATGCTGAGGGGATCTTGTACGACAGATCGATCTGAGCAGATTTCTGCTG 858
OY 859 GCCAAGGCGCGCGCGGATCCGCGCGGCACTATTACGCGCGCGCGCAAGCGCGCTGCTG 918
    |||||
DB 859 GCGAAGCTGAGCCGATCCGCTCGGCGCACTATTGCGCGCGCTGACGTGCTGCTGCTCAG 918
OY 919 CTGATCGACG 929
    |||||
DB 919 TTCGACCGCG 929
```

## RESULT 4

```
US-10-919-182-11
; Sequence 11, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-11
```

```
Query Match      27.3%; Score 283.8; DB 9; Length 1110;
Best Local Similarity 57.0%; Pred. No. 6.1e-62;
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;
```

```
OY 19 AAGTATCGCGCGCGCGGTGACAGCGCGCGCGGTGTTCTCGATCTGACCGCACAGTC 78
    |||||
DB 19 AAGTCTCTCGCGCAACCGTTACAGGACAGCCGGTATGGCTCGACGACGACGCAACGATC 78
OY 79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGACGTCGCGCTGATCGCATTC 138
    |||||
DB 79 GACAAAGTCGATCGGCATCATCGAAGAGCTGCCCAAAAGGCGCGAGTCTGATCGCTTTC 138
OY 139 CCAGAGACTTGATTCGCGCTATCCCTTTGATATGCTGGCGCGCGCGCTTGAGGC 198
    |||||
DB 139 CCGAAGTATTCAATCCGGGCTACCCCTATTGGCGGTGCTCGGCGACGTGAAGTACAGC 198
OY 199 ATGCGCTTCGTCACGCGCTATTTGAGAATTGCTGCTGGCGGCAAGCAGTGGCAG 258
    |||||
DB 199 CTAAGCTTACTTCACGCTATCACGAAATTCGTTGAGCTAGGTGACGACCGTATGCCGT 258
OY 259 GCCCTGGCGGATGCGGCGCGCGCGCGCGCATGATGTCGTGGCGCGCTATAGCGAGCGC 318
    |||||
DB 259 CGCCTCAAGTGGCGCGCGCGCGCAACAAATCGCACTGTCATGGGCTATTGCGAGCGG 318
OY 319 GCGGCGGCGAGCCTTATATGCGCGCGCGCATCTTCGCGCGCGCGATGGCGATCTGATCGCC 378
    |||||
```

Db 319 GAAGCCGATCGCGTATCTGAGCCAGGTGTTCAATCGACGAGCGTGCGGAGATCGTTGCC 378  
QY 379 GCGCGCCGAAGCTCAAGCCTACCCATGCGAGCGCACCGTGTCGCGGAGAGACGGC 438  
Db 379 AATCGCGCAAGCTGAAGCCACACACGTTGAGCGTACGATCTACGGCGAAGCAACGGA 438  
QY 439 AGCCATCTCGCGGTGACGATACCGCCATCGGGCCCTCGGGCGCTCTGTTGCTGGAG 498  
Db 439 ACCGATTTCTCAGCGACGACTTCGCGTTCGACCGCGTGGTGATTGAATGCTGGAA 498  
QY 499 CACATCCAGCCATTGTCGAATAACGCCATGTACGCCCGCGACGAACAGTTCACGTGCG 558  
Db 499 CATTTCCAACCGCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558  
QY 559 TCGTGCCGAGCTTCAGCCTCTATCGCGGCATGGCCTATGCGCTCGGACCGAGGTCAAT 618  
Db 559 TCGTGCCGCGGATGTCCCTCTTCAAGCCGATGTTTCCAACTGAGATCGAAAGCCAAAC 618  
QY 619 ACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTACGTGCTGGCGTGCAGCG 678  
Db 619 GCGACGGTCTGCGCTGCTGACGCAATCGAAGGCCAAACCTTTGTGCTTGTCTGACGAG 678  
QY 679 ACCGTTTCGCGGAGATGATCAAGTATTGTGATACGCCCGACGAAGAGATGTTCTC 738  
Db 679 GTGATCGAGCTAGCGCGCATCGAAACGTTCTGCTCAACGACGAACGAGCGCACTGTTG 738  
QY 739 AAGCGCGCGCGGTTTTCATGATTTTCGAGCCCGACGCGCGCGCTGGCCGAGCCG 798  
Db 739 CCGCAAGATGTGGCTGGCGCGCATTTACGCGCCCGATGGAAGCAGCTTGCGAAGCCT 798  
QY 799 CTCCCGAGACCGAAGAGGAGTGTGTGCGCGATATCGACCTCGGATGATCGCGTTG 858  
Db 799 CTGCGGAAGATGCTGAGGGGATCTGTACGAGAGATCGATCTGAGCAGATTCTGCTG 858  
QY 859 GCCAAGCGCGCGCGATCCGCGCGGCACTATTCAAGGCCGACGTAACGCGGCTGCTG 918  
Db 859 GCGAAGGCTGAGCGCGATCCGCTCGGGCACTATTGCGCGCTGACGTGCTGCTGCCAG 918  
QY 919 CTGATCGACG 929  
Db 919 TTCGACCCGCG 929

RESULT 5  
US-10-919-182-5  
; Sequence 5, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: C12584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase B2 and H9  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-5

Query Match 27.1%; Score 282.2; DB 9; Length 1110;  
Best Local Similarity 56.9%; Pred. No. 1.5e-61;  
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;

QY 19 AAGTATCGGGCGCGCGGTGACAGGCCGCGCGGTGTTCTTCGATCTCGACCGCACAGTC 78  
Db 19 AAGTCTCTCGCGCAACCGTTACAGGAGAGCCGGTATGGCTCGACGCAAGCAACGATC 78  
QY 79 GAGAAAGCGATCGGCCCTGATCGAGCAGCGCGCAAGCAGAGAGTGGCTGATCGCATTC 138  
Db 79 GACAAAGTATCGGCATCATCGAAGAGCTGCCCAAAAGGCGCGAGTCTGATCGCTTTC 138  
QY 139 CCAGAGCTTGATTTCCCGCTATCCCTTTTGGATATGGCTGGCGCGCGCGCTTGGGGC 198  
Db 139 CCGAAGTATTTACTTCAGCTATCGAAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGT 258  
QY 199 ATGCGCTTCGTCAGCGCTATTTGAGAATTGCTCGTGGCGCGGCAAGCAGTGGCAG 258  
Db 199 CTAAGCTTTACTTCACGCTATCGAAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGT 258  
QY 259 GCCGTGGCGGATGCGCGCCGCGCAAGCATGATGTGCGCCGCGCTATAGCGAGCGC 318  
Db 259 CGCCTCAGCTGGCGCGCGCGCAACAATCGCACTCGTCATGGGCTATTCCGAGCGG 318  
QY 319 GCGGCGGCAAGCCTTATATGGGCCAGCGCATCTTCGCGCCCGATGGCGATCTGATCGCC 378  
Db 319 GAAGCCGATCGCGCTATCTGAGCAGAGGTTCATCGACGAGCGTGGCAGATCGTTGCC 378  
QY 379 GCGCGCGCAAGCTCAAGCCTTATGCGGAGCGCACCGCTGTTCCGCGAGGAGACGGC 438  
Db 379 AATCGCGCAAGCTGAAGCCACACACGTTGAGCGGTACGATCTACGGCGAAGGCAACGGA 438  
QY 439 AGCCATCTCGCGGTGACGAGATCCCATCGGGCGCTCGCGCGCTCTGTTGCTGGAG 498  
Db 439 ACCGATTTCTCAGCAGCAGCTTCGCGTTGGAACGCGGTGGATTGAATGCTGGGAA 498  
QY 499 CACATCCAGCCATTGTGCAATAACGCCATGTACGCCCGCGCAAGCAGGTCCACGTGCG 558  
Db 499 CATTTCCAACCGCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558  
QY 559 TCGTGCCGAGCTTCAGCCTCTATCGCGGCATGGCCTATGCGCTCGGACCGAGGTCAAT 618  
Db 559 TCGTGCCGCGGATGTCCCTCTTCAAGCCGATGTTTCCAATGAGCATGAAAGCCAAAC 618  
QY 619 ACCGCCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTACGTGCTGCGTGGCGG 678  
Db 619 GCGACGGTCCCGCTGCTGACGCAATCGAAGGCCAAACCTTTGTGCTTGTCTGACGAG 678  
QY 679 ACCGTTTCGCGGAGATGATCAAGTATTGTGATACGCCCGACGAAGAGATGTTCTC 738  
Db 679 GTGATCGAGCTAGCGCGATCGAAACGTTCTGCTCAACGACGAACGCGCACTGTTG 738  
QY 739 AAGCGCGCGCGGTTTTCATGATTTTCGCGCGGCGCGCGCGCGCTGGCGAGCCG 798  
Db 739 CCGCAAGATGTGGCTGGCGCGCATTTACGGCCCGATGGAAGCAGCTTGCGAAGCCT 798  
QY 799 CTCCCGAGACCGAAGAGGAGTGTGTGCGCGATATCGACCTCGGATGATCGCGTTG 858  
Db 799 CTGCGGAAGATGCTGAGGGGATCTGTACGCAAGATCGATCTGAGCAGATTCTGCTG 858  
QY 859 GCCAAGCGCGCGCGATCCGCGCGGCACTATTCAAGGCCGACGTAACGCGGCTGCTG 918  
Db 859 GCGAAGGCTGAGCGCGATCCGCTCGGGCACTATTGCGCGCTGACGTGCTGCTGCCAG 918  
QY 919 CTGATCGACG 929  
Db 919 TTCGACCCGCG 929

RESULT 6  
US-10-919-182-13  
; Sequence 13, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark



```
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
; US-10-919-182-13
```

```
Query Match      27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;
```

```
QY 19 AAGTATCGCGCGCGCGGTGCAGGCGCGCGGTTCCTCGATCTGCACCGCACAGTC 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19 AAGTCTCTCGCGCAACCGTTCAGGCGAGCGCGGTATGGCTCGACCGCAGCAACGATC 78

QY 79 GAGAAAGCGATCGCGCTGTATCGAGCAGCGCGCAAGCAGACGTGCGCTGATCGCATTC 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 GACAAATCGATCGGCATCATCGAAGAGCTGCCAAAGGGCGCGAGTCTGATCGCTTTC 138

QY 139 CCAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCTGGCGCGCGCGCTTGGGCG 198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 CCGAAGTATTCATTCGGGCTACCCCTATTTGGCGCTGGCTCGGCGAGCTGAAGTACAGC 198

QY 199 ATGCGCTTCGTCAGCGCTATTTGAGAATTGCTGCTGCGCGCAGCAAGCAGTGGCAG 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 CTAAGCTTTACTTTCACGCTATCAGAGAAATTCGTTGAGCTAGTGAGCAACCGTATGCGT 258

QY 259 GCCCTGGCGGATCGCGCGCGCGCGCGCGCGCATGCTGCGCGCGCTATAGCGAGCGC 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 CGCCTCCAGCTGCGCGCGCGCGCGCGCAACAAATTCGCACTGTCATGGGCTTATTCGAGCGG 318

QY 319 GCGGCGCGCAGCTCTATATGCGCAGCGCATCTTCGCGCGCGCGATGGCATCTGATCGCC 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 GAAGCCGATCGCGCTATCTGAGCCAGGTGTTTCATCGACGAGCGTGGAGATCGTTGCC 378

QY 379 GCGCGCGCAGCTCAAGCCTTACCATGCGAGCGCAGCCGTTGTCGCGAGAGACGCGC 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 AATCGCGCAGCTGAAGCCACACACAGTTGAGCTGACATCTACGCGCAAGCAACGGA 438

QY 439 AGCCATCTCGCGGTGCAAGATACCGCATCGCGCGCGCTCGCGCGCTGTGCTGGAG 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 439 ACCGATTTCTCAGCAGCAAGCTTCGCGTGGACGCGTGGATGAACTGCTGGGAA 498

QY 499 CACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCAGCAAGTCCACGTGCGC 558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 499 CATTAACAACCCCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558

QY 559 TCGTGCGCGAGCTTCAGCCTCTATCGCGGCGATGGCCTATGCGCTCGGACCGGAGGTCAAT 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 559 TCGTGCGCGGAGTGTCCCTCTTACGCCGAGTGTTCACACTGAGCATCGAAAGCCAAC 618

QY 619 ACCGCGCAGCAGATCTACGCGGTGAGGCGGCTGCTACGTGCTGGCGGTGCGCG 678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 619 GCGACGTCACCCGCTCTGACGCAATCGAAGCCAAACCTTTGTGCTTGTGCTGACGCGAG 678

QY 679 ACCGTTTCGCGGAGATGATCAAGTATTGTGATACGCCCGAAGAGAGATGTTCTC 738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 679 GTGATCGAGACCTAGCGCGATCGAAACGTTCTGCTCAACGACGAAACAGCGCACTGTTG 738

QY 739 AAGCGCGCGCGGTTTTTGGCATGATTTTCGCGCGCGCAGCGCGCGCTGGCGAGCGC 798
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 739 CCGCAAGGATGTGCTGGCGCGCATTTACGCGCGCGGATGGAAGCGAGCTTGGCAAGCCT 798
```

```
QY 799 CTCCCGAGACCCGAAGAGGACTGCTGTGTCGCCGATATCGACCTTGGCATGATCGCGTGTG 858
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 799 CTGGCGGAGATGCTGAGGGGATCTTGTACGACAGATCGATCTGGACAGATTTCTGCTG 858

QY 859 GCCAAGCGCGCGCGCGATCCGCGCGGCACTATTACGCGCGCGCAGTAACGCGGCTGCTG 918
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 859 GCGAAGCTGAGCCGATCCGCTGCGGCACTATTGCGCGCGCTGACGTGCTGTGCTGTCAG 918

QY 919 CTGATCGACG 929
    ||||| |||||
DB 919 TTCGACCGCG 929
```

```
RESULT 7
US-10-919-182-15
; Sequence 15, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
; US-10-919-182-15
```

```
Query Match      27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;
```

```
QY 19 AAGTATCGCGCGCGCGGTGCAGGCGCGCGGTTCCTCGATCTGCACCGCACAGTC 78
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 19 AAGTCTCTCGCGCAACCGTTCAGGCGAGCGCGGTATGGCTCGACCGCAGCAACGATC 78

QY 79 GAGAAAGCGATCGCGCTGTATCGAGCAGCGCGCAAGCAGACGTGCGCTGATCGCATTC 138
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 79 GACAAATCGATCGGCATCATCGAAGAGCTGCCAAAGGGCGCGAGTCTGATCGCTTTC 138

QY 139 CCAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCTGGCGCGCGCGCTTGGGCG 198
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 139 CCGAAGTATTCATTCGGGCTACCCCTATTTGGCGCTGGCTCGGCGACGTGAAGTACAGC 198

QY 199 ATGCGCTTCGTCAGCGCTATTTGAGAATTGCTGCTGCGCGCGCAGCAAGCAGTGGCAG 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 199 CTAAGCTTTACTTTCACGCTATCAGAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGT 258

QY 259 GCCCTGGCGGATCGCGCGCGCGCGCGCGCGCATGCTGCTGCGCGCGCTATAGCGAGCGC 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 259 CGCCTCCAGCTGCGCGCGCGCGCGCAACAAATTCGCACTGTCATGGGCTATTCGAGCGG 318

QY 319 GCGGCGCAGCCTCTATATGCGCAGCGCATCTTCGCGCGCGCGATGCGATCTGATCGCC 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 319 GAAGCCGATTCGCGCTATCTGAGCCAGGTGTTTCATCGACGAGCGGTGGCAGATGTTGCC 378

QY 379 GCGCGCGCAAGCTCAAGCCTTACCATGCGAGCGCAGCCGTTGTCGCGAGAGAGCGCGC 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 379 AATCGCGCAGCTGAAGCCACACACAGTTGAGCTGACGATCTACGCGCAAGCAACGGA 438

QY 439 AGCCATCTCGCGGTGCAAGATACCGCATCGCGCGCGCTCGCGCGCTGTGCTGGAG 498
```



```

Db 439 ACCGATTTCTCAAGCAGACTTCGCGTTCGGAACGGTGGATTTGAATGCTGGAA 498
QY 499 CACATCCAGCCATTGTGGAATACGCCATGTACGCCCGGACGAGCAAGTCCACGTCCG 558
Db 499 CATGTTCAACCGCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558
QY 559 TCGTGCCGAGCTTCAGCCTCTATCGCCGCGCATGGCCTATGCGCTCGGACCGAGGTCAAT 618
Db 559 TCGTGCCGCGGATGTCCCTCTTACAGCCGAGTGTTCACACTGAGCATCGAAAGCCAAAC 618
QY 619 ACCGCCCAAGCCAGATCTACGCGGTGAGGCGCGCTCTACGCTGCGGTGCGCGG 678
Db 619 GCGACGGTCAACCGCTCTGTAACGAATCGAAGGCCAAACCTTGTGCTTGTGCTGACGCGAG 678
QY 679 ACCGTTCCGCGGAGATGATCAAGGTATTGTGGATACGCCCGACAGAGAGATGTTCTTC 738
Db 679 GTGATCGGACCTTAGCGCGATCGAAACGTTCTGCTCAACGACGAAACGCGGCACTGTTG 738
QY 739 AAGCGCGCGCGGCTTTTCCATGATTTTCGAGCCCGACGCGCGCGCTGCGGAGCCG 798
Db 739 CCGCAAGGATGTGCTGGCGCGCATTTACGCGCCGAGTGAAGCGAGCTTGCGAAGCCT 798
QY 799 CTCCCGGAGACGAAGAGGACTGTGTGCGCGCATATCGACCTCGGCATGATCGCGTTG 858
Db 799 CTGCGGAAGATGTGAGGAGATCTGTACGAGAGATCGATCTGAGCAGATTTCTGCTG 858
QY 859 GCCAAGCGCGCGCGCATCCGCGCGGCACTATTCACGCGCCGACGTAACGCGGCTGCTG 918
Db 859 GCGAAGGCTGAGCGCGATCCGCTCGGCGCACTATTGCGCGCTGACGCTGTGCTGCCAG 918
QY 919 CTGATCGACG 929
Db 919 TTCGACCCGCG 929

```

RESULT 8

```

US-10-919-182-17
; Sequence 17, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: C12584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-17

```

```

Query Match 27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;
QY 19 AAGTATCGCGCGCGGTGTCAGGCGCGCGGTGTTCTGATCTGACCGCACAGTC 78
Db 19 AAGTTCCTCGCGCAACCGTTCAAGCAGAGCCGGTATGCTCGACGACGCAACGATC 78
QY 79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGCAAGCAGAGACGTGCGCTGATCGCATTC 138
Db 79 GACAAAGTCATCGGCATCATCGAAGAACTGCCCAAAAGGCGGAGTCTGATCGCTTTC 138

```

```

QY 139 CCAGAGACTTGGATTCCCGGCTATCCCTTTTGATATGCTGGCGCGCGCTTGGGGC 198
Db 139 CCGGAAGTATTCTATTCGGGGCTACCCCTATTGGCGGTGGCTCGGCGAGCTGAAGTACAGC 198
QY 199 ATGCGCTTCCTCCAGCGCTATTTCAGAAATTCGCTCGTGCGCGGACGAAGCAGTGGCAG 258
Db 199 CTAACTTACTTCAACGCTATCAAGAAATTCGTTGAGCTTAAGTGAACCGTATGCGT 258
QY 259 GCGGCGGAGTCCGCGCGCGCGCGCGCATGATGTCGTGCGCGCTATAGCGAGCGC 318
Db 259 GCGCTCAAGCTGCGCGCGCGCGCGCAAAATCGCACTTCATGAGGCTATTTCGAGCGG 318
QY 319 GCGGCGGAGCTCTATATGCGCGAGCATCTTCGCGCGCGCGCATGCGCGCGATCTGATCGCC 378
Db 319 GAAGCCGATTCGCGCTATCTGAGCGAGTGTTCATCGACGAGCGTGGCGAGATCGTTGCC 378
QY 379 GCGCGCGCAAGCTCAAGCCTACCAATGCGGAGCGCACCGGTTCGGCGAGGAGACGCGC 438
Db 379 AATCGCGCAAGCTGAAGCCACACAGCTTGAAGCGTACGATCTACGCGCAAGCAACGCA 438
QY 439 AGCCATCTCGCGGTGACGATACCGGCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAG 498
Db 439 ACCGATTTCTCACGACGACTTCGCGGTTCGAGCGCGCTCGGTGATTTGAATGCTGGAA 498
QY 499 CACATCCAGCCATTGTGGAATACGCCATGTACGCCCGCGCAACAGTCCACGTCCGCG 558
Db 499 CATCTACAACCGCTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCA 558
QY 559 TCGTGCGGAGCTTCAAGCCTCTATTCGCGGATGCGCTATGCGCTCGGAGGTCAAT 618
Db 559 TCGTGCGCGGAGTGTCCCTCTTCAAGCCGATGTTTCCAACTGAAGCATCGAAAGCCAAAC 618
QY 619 ACCGCCCAAGCCAGATCTACGCGGTGAGGCGCGGCTGCTACGCTGCGCTGCGCGC 678
Db 619 GCGACGGTCAACCGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGCTGACGCGAG 678
QY 679 ACCGTTCCGCGGAGATGATCAAGTATTGTGATACGCCCGCAACAGAGATGTTCTTC 738
Db 679 GTGATCGGACCTTAGCGCGATCGAAACGTTCTGCTCAACGACGAAACAGCGGCACTGTTG 738
QY 739 AAGCGCGCGCGGCTTTTGCATGATTTTCGCGCGCGCGCGCGCGCTGCGCGGAGCGC 798
Db 739 CCGCAAGGATGTGCTGGCGCGCATTTACGCGCGCGGATGGAAGCGAGCTTGCGAAGCCT 798
QY 799 CTCCCGGAGACCGAAGAGGACTGCTGTCGCGCGATATCGACCTCGGCATGATCGCGTTG 858
Db 799 CTGCGGAAGATGTGAGGAGATCTTGTACGAGAGATCGATCTGAGCAGATTTCTGCTG 858
QY 859 GCCAAGCGCGCGCGCATCCGCGCGGCACTATTTCAGCGCGCGCAAGTACGCGGCTGCTG 918
Db 859 GCGAAGGCTGAGCGCGATCCGCTCGGCGCACTATTTCGCGCGCTGACGTGCTGCTGCCAG 918
QY 919 CTGATCGACG 929
Db 919 TTCGACCCGCG 929

```

RESULT 9

```

US-10-496-351-4
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wei
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; FILE REFERENCE: UMAS-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0

```

```
; SEQ ID NO 4
; LENGTH: 9975
; TYPE: DNA
; ORGANISM: Actinosynema pretiosum
US-10-496-351-4
```

```
Query Match      4.9%; Score 50.6; DB 7; Length 9975;
Best Local Similarity 43.7%; Pred. No. 0.0025;
Matches 325; Conservative 0; Mismatches 409; Indels 9; Gaps 2;
```

```
QY 186 GCCGCTTGGGCGATGCGCTTCGTCACGCGCTATTTCGAGAATTGCTCGTCCGCGGCAG 245
    |||||
DB 3588 GCCGGTGTGCGCCACGACGTCCCGCGGCTCCGCGAGGACGGGTGTCTCACCT 3647
QY 246 CAAGCAGTGGCAGGCGCTGGCGGATGCGCCCGCCACGCGCATGTCATGTCGCGCG 305
    |||||
DB 3648 CGCGGTGTGACAGGGTTCCTCGCCGATCCCGGACGCGGACTCGGTGCTGCTGTGT 3707
QY 306 CTATAGCAGCGCGCGCGGCGCAGCCTCTATATGGGCCAGCGATCTTCGGCCCGATGG 365
    |||||
DB 3708 CACCAAGGGCGGCGAGGGCCCGGACGCGCGCAGCCGCGGACGTCCGCGTGCACGGCT 3767
QY 366 CGATCTGATCGCGCGCGCCGCAAGCTCAAGCCTACCATGCGGACGCACCGTGTTGG 425
    |||||
DB 3768 GTCCGCTCCGCCACGAGCGGAGGAGCGCGCGCGGATGCTCTGTGGACACCGACGCTC 3827
QY 426 CGAGGAGACGCGCAGCATCTCGCGGTGACGATACCGCATCGGGCGCTCGCGCGCT 485
    |||||
DB 3828 CGCCCCGACGACTGATCAACGCGCTCGCGCGGACCGAGCCCCGACGCTGCGCTGCG 3887
QY 486 CTGTGCTGGGAGCACATCCAGCCATTGTGAAATACG---CCATGTACGCGCGCAGCA 542
    |||||
DB 3888 CGCGGCGCGGTCAACGCGCGGAGGTTCGCGCGCGCGCGCGCGCGCGCGCACCGCT 3947
QY 543 ACAGGTCCACGTGCGCTGTGGCGGAGCTTCAGCCTCTATCGCGCATGCGCTATGCGCT 602
    |||||
DB 3948 CTTGACCCCGAGGGCACCGTCTGTGTACCGCGGACACCGCGTGTGCTCGCGCGCGCT 4007
QY 603 CGGACCGGAGGTCAATACCGCGCGCAAGCCAGATTAACCGGTGAGGGCGGCTGTAAGT 662
    |||||
DB 4008 CGCCCGCACCTGCTCAACGCGCGCGCGCGCGCGCGCACCTGCTGTCTGACCGC----- 4062
QY 663 GCTGCGCTCGTCCGCGACCGTTTCGCGCGAGATGATCAAGGTATTGTTGATACGCCCGA 722
    |||||
DB 4063 -CGGGGCCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 4121
QY 723 CAAGGAGATGTTCTCAAGCGCGCGCGGCTTTGTCATGATTTTCGCGCGCGACGCGCG 782
    |||||
DB 4122 CCGCGTGTGTGTCGCGCGCTGACGTGCGCGACCGCGCGCGCTCGCGCGCGCTCGA 4181
QY 783 CGCCCTGCGCGAGCGCTCCCGGAGACCGAAGGAGACTGCTGTGCGCGATATCGACCT 842
    |||||
DB 4182 CGCGGTCCCGCGCGGACCGCGTCAACCGCGCGGTGTGACCTGCGCGCGCTCGACGA 4241
QY 843 CGGCATGATCGCGTTGGCCCAAGCGCGCGCGCGATCCGCGGGGCACTATTCAACGCCCGA 902
    |||||
DB 4242 CGGACGATCACCGCGCTCAACCCCGACCGCGGTGCGCGGGTCTCGCCCCCAAGGCCAC 4301
QY 903 CGTAACGCGGCTGCTGTGATC 925
    |||||
DB 4302 CGCGCGCGCGCACCTGACGAGC 4324
    |||||
```

## RESULT 10

```
US-10-496-351-56/c
; Sequence 56, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
```

```
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinosynema pretiosum
US-10-496-351-56
```

```
Query Match      4.9%; Score 50.6; DB 7; Length 82746;
Best Local Similarity 43.7%; Pred. No. 0.0023;
Matches 325; Conservative 0; Mismatches 409; Indels 9; Gaps 2;
```

```
QY 186 GCCGCTTGGGCGATGCGCTTCGTCACGCGCTATTTCGAGAATTGCTCGTCCGCGGCAG 245
    |||||
DB 63613 GCCGGTGTGCGCCACGACGTCCCGCGGCTCCGCGAGGACGGGTGTCTCACCT 63554
QY 246 CAAGCAGTGGCAGGCGCTGGCGGATGCGCCCGCCACGCGCATGTCATGTCGCGCG 305
    |||||
DB 63553 CGCGGTGTGACAGGGTTCCTCGCCGATCCCGGACGCGGACTCGGTGCTGCTGTGT 63494
QY 306 CTATAGCAGCGCGCGCGGCGCAGCCTCTATATGGGCCAGCGATCTTCGCGCGCGATGG 365
    |||||
DB 63493 CACCAAGGGCGGAGGGCCCGGACGCGCGGACCGCGGACTCGCGCGTGCACGGCT 63434
QY 366 CGATCTGATCGCGCGCGCCGCAAGCTCAAGCCTACCATGCGGACGCAACGTGTTGG 425
    |||||
DB 63433 GTCCGCTCCGCCACGAGCGGAGGAGCGCGCGGATGCTCTGTGGACACCGACGGCTC 63374
QY 426 CGAGGAGACGCGCAGCATCTCGCGGTGACGATACCGCATCGGGCGCTCGCGCGCT 485
    |||||
DB 63373 CGCCCCGACGACTGATCAACGCGCTCGCGCGCGGACCGAGGCCACGCTGCGCTGCG 63314
QY 486 CTGTGCTGGGAGCACATCCAGCCATTGTGAAATACG---CCATGTACGCGCGCAGCA 542
    |||||
DB 63313 CGCGGCGCGGTCAACGCGCGGAGTTGCGCGCGCGCGCGCGCGCGCGCACCGCT 63254
QY 543 ACAGGTCCACGTGCGCTGTGGCGGAGCTTCAGCCTCTATCGCGCGCATGCGCTATGCGCT 602
    |||||
DB 63253 CTTGACCCCGAGGGCACCGTCTGTGTACCGCGGACACCGGCTGTGCGCGCGCT 63194
QY 603 CGGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGGCGGCTGTAAGT 662
    |||||
DB 63193 CGCCCGCACCTGCTCAACGCGCGCGCGCGCGCGCGCACCTGCTGTCTGACCGC----- 63139
QY 663 GCTGCGCTCGTCCGCGACCGTTTCGCGCGAGATGATCAAGGTATTGTTGATACGCCCGA 722
    |||||
DB 63138 -CGGGGCCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 63080
QY 723 CAAGGAGATGTTCTCAAGCGCGCGCGGCTTTGTCATGATTTTCGCGCGCGACGCGCG 782
    |||||
DB 63079 CCGGTGTGTGTCGCGCGCTGACGTGCGCGACCGCGCGCGCTCGCGCGCGCTCGA 63020
QY 783 CGCCCTGCGCGAGCGCTCCCGGAGACCGAAGGAGACTGCTGTGCGCGATATCGACCT 842
    |||||
DB 63019 CGCGGTCCCGCGCGGACCGCGTCAACCGCGGTGTGACCTGCGCGCGCGCTCGACGA 62960
QY 843 CGGCATGATCGCGTTGGCCCAAGCGCGCGCGCATCCGCGGGGCACTATTCAACGCCCGA 902
    |||||
DB 62959 CGGACGATCACCGCGCTCAACCCCGACCGCGGTGCGCGGGTCTCGCCCCCAAGGCCAC 62900
QY 903 CGTAACGCGGCTGCTGTGATC 925
    |||||
DB 62899 CGCGCGCGCGCACCTGACGAGC 62877
    |||||
```

## RESULT 11

```
US-11-096-568A-20685
; Sequence 20685, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
```

```

; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theryby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20685
; LENGTH: 1706
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1706)
; OTHER INFORMATION: Ceres Seq. ID no. 12386684
US-11-096-568A-20685

```

Query Match	4.6%;	Score 48.2;	DB 11;	Length 1706;
Best Local Similarity	49.8%;	Pred. No. 0.011;		
Matches 122; Conservative	0;	Mismatches 123;	Indels 0;	Gaps 0;

[illegible]

```

RESULT 12
US-10-478-943D-1
; Sequence 1, Application US/10478943D
; Publication No. US20060073574A1
; GENERAL INFORMATION:
; APPLICANT: University College Dublin, National University of Ireland
; TITLE OF INVENTION: Engineered biosyntheses of novel polyenes
; FILE REFERENCE: PC-1636US
; CURRENT APPLICATION NUMBER: US/10/478, 943D
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: S2001/0527
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 113193
; TYPE: DNA
; ORGANISM: Streptomyces nodosus
US-10-478-943D-1

```

Query Match	4.6%	Score 47.8;	DB 7;	Length 113193;
Best Local Similarity	49.4%;	Pred. No. 0.011;		
Matches 124; Conservative	0;	Mismatches 127;	Indels 0;	Gaps 0;

QY	788	TGGCCGAGCCGCTCCCGGAGACCGAAGAGGA	CTGCTGTGC	CCGATATCGACTTCGGCA	847
Dd	46971	TGGCGGGGGTGTCTCGCTGGAGGACGCTTGCCGGCTG	GTCGCCCGCCCGCGCGCTCA		47030
QY	848	TGATCGCGTTGGCCAAAGCGCGCGCCGATCCGGCGGC	CACTATTTCACGCGCCGACGTAA	907	
Dd	47031	TGGACGCGCTCCCCCGGGCGCGCCCATGGCCCGCTGAGG	CACCGAGACGAGGTGC	47090	

[illegible]

```

RESULT 13
US-10-496-351-8
; Sequence 8, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wein
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2247
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-8

```

Query Match	4.6%	Score 47.6	DB 7	Length 2247
Best Local Similarity	45.6%	Pred. 0.015		
Matches 167	Conservative 0	Mismatches 199	Indels 0	Gaps 0
QY	201	GCGETTCGTCACGCGCTATTTCGAGAATTGCTCGTGGCGGGCAGCAAGCAGTGGCAGGC	260	
Db	1527	GCTGCACCGCGTGTCTTCTCTCTACGAGGGCCGCCGCGGTGCTGACCGGTTGTGCT	1586	
QY	261	CCTGGCGGATGCGGCCCGCCCGCACGGCATGCAATGCTGGCCGGCTATAGCGAGCGCG	320	
Db	1587	GCGCGCGGCGGGCGGGGTACCGGCATGTCGGCGGACCGGCTCGGCAAGACCAC	1646	
QY	321	GGCGCGCAGCCTTATATGGGCCAGGCGATCTTGGCCCCGATGGCATCTGATCGCGC	380	
Db	1647	CGTCGCCAAGCTGTGATGCGCTTCCAGAGACGCGGACACCGCCGGGTGCTCGACGG	1706	
QY	381	GCGCCGCAAGCTCAAGCCTAACCCATGCGAGCGCACCGTGTTCGGCAGGAGACGGCAG	440	
Db	1707	GCGGACGTGCGCGGCTGCGCCTGCCCGACCTGCGCGCGGCCGTGGGTTCTGCGCCA	1766	
QY	441	CCATCTCGCGGTGACGATACCGCCATCGGGCGCTCGGCGCTGTGTGCTGGAGCA	500	
Db	1767	GGACCCGTTCTGTTCGACGGCAGATCGCGACACATCCGCTACGGCAGTTCTCCGC	1826	
QY	501	CATCCAGCCATTGTGAAATACGCCCATGTACGCCCGCAGCAACAGSTCCACGTCCGCTC	560	
Db	1827	CACCGACGACGAGGTCCCGGAGGCCCGCCGACGGCCAGGCCACCGCTTCTGCAGTTC	1886	
QY	561	GTGGCC 566		
Db	1887	GCTGCC 1892		

RESULT 14  
US-10-496-351-3  
; Sequence 3, Application US/10496351  
; Publication No. US20060084141A1  
; GENERAL INFORMATION:

```

; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wei
; APPLICANT: Leistner, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; TITLE OF INVENTION: Ansamitocin
; FILE REFERENCE: UWASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Actinosynnema pretiosum
US-10-496-351-3

```

Query Match	4.6%	Score 47.6	DB 7	Length 4770
Best Local Similarity	49.2%	Pred. No. 0.015		
Matches 125	Conservative 0	Mismatches 129	Indels 0	Gaps 0
QY	232	CTCGTCGCGCGGAGCAAGCAGTGGCAGGCCCTGTGGCGATGCGGCCGCCAGCGGCATG	291	
DB	3649	CTGGAGCGCGCGGCGAGCAGGTCACCGCCGTGTGACCGCGCGGCCCAACGCCAG	3708	
QY	292	CATGTCGTGGCGCGCTATAGCAGCGCGCGCGCGCAGCCTCTATATGGCCAGCGCATC	351	
DB	3709	ACCCCGTCGCCGACACACACCCCGAGGAGCAGCGCGCGTCCAGCCGCCAAGCGGCTC	3768	
QY	352	TTGGCCCCGATGCGATCTGATCGCCGCGCGCCGCAAGCTCAAGCCTACCATGGCAG	411	
DB	3769	GGCGCCGAGCACCTCGACGAGCTGTGGCGGGCGGCCGCTCGACGCGTTCTGTCCTGTC	3828	
QY	412	CGCACCGGTGTTCCGCGGAGGAGACGGCAGCCATCTCGCGGTGCACGATACCGCCATCGG	471	
DB	3829	TCCTCCAACGCGGGCGGTGTGGGGCAGCGGCGCCAGTGCAGCTACGCCCGCCCAACGCC	3888	
QY	472	CGCCTCGGCGCGCT	485	
DB	3889	CGCCTCGACGCACT	3902	

RESULT 15  
US-10-933-746-36  
; Sequence 36, Application US/10933746  
; Publication No. US2006002998A1  
; GENERAL INFORMATION:  
; APPLICANT: FLECKENSTEIN, Bernhard  
; ENSER, Armin  
; TITLE OF INVENTION: HUMAN SEMAPHORIN L (H-SEMA) AND  
; CORRESPONDING SEMAPHORINS IN OTHER SPECIES  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Frommer Lawrence & Haug LLP  
; STREET: 745 Fifth Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10151  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/933,746  
; FILING DATE: 03-Sep-2004  
; CLASSIFICATION: <Unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lawrence, William F.  
; REGISTRATION NUMBER: 28,029  
; REFERENCE/DOCKET NUMBER: 514429-3647  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-588-0800

```

; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 8192 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 36:
US-10-933-746-36

```

Query Match	4.6%;	Score 47.6;	DB 9;	Length 8192;
Best Local Similarity	45.3%;	Pred. No. 0.014;		
Matches 173;	Conservative 0;	Mismatches 209;	Indels 0;	Gaps 0;
QY 177	GCTGGGCGCGCGCTTTGGGGCATGCGCTTCGTCCAGCGCTATTTCGAAGATTCCGTCGT	236		
Db 2932	GCTGCCACACTCACTCTTGCGCTTGCTGTCACATGATGAGCAAGGGCGAGGAGCTGTT	2991		
QY 237	GCGCGGCGAGCAAGCATGGCGAGGCGCCCTGGCGGATGCGGCGCGCCGACCGCATGCATGT	296		
Db 2992	CACCGGGGTGGTCCCATCTGTGTCGAGCTGACGGCGACGTAAACGGCCACAAGTTCAAG	3051		
QY 297	CGTGCCCGCTATAGCAGCGCGCGGCGCGGCGGCGAGCCTTATATGGCCAGGCGATCTTCG	356		
Db 3052	CGTGCCCGCGAGGGCGAGGGCGATGCCACTACGSCAAAGCTGACCTGAAGTTCAATCG	3111		
QY 357	CCCCGATGGCATCTGATCGCGCGCGCGCAAGCTCAAGCTTACCCTAGCGGAGCGCAC	416		
Db 3112	CACCACCGGCAAGCTGCCGTCCTGCGCCACCTCGTGACCAACCTGACCTACGGCGT	3171		
QY 417	CGTGTTCGGGAGGAGACGGCGAGCCATCTCGCGGTGCACGATACCGCATCGGGCGCCT	476		
Db 3172	GCAGTGCTTCAGCCGCTACCCCGACCATGAGCAGACGACTTCTCAAGTCCGCCAT	3231		
QY 477	CGCGCGCTCTGTGCTGGGAGCACATCCAGCCATTGTGGAATACGCCATGTACGCCGC	536		
Db 3232	GCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGAGCAGCGCACTACAGAC	3291		
QY 537	CGACGAACAGTCCAAGTCCGCG 558			
Db 3292	CCGCGCCGAGGTGAAGTTCGAG 3313			

Search completed: April 27, 2006, 00:22:19  
Job time : 630.12 secs



**THIS PAGE BLANK (USPTO)**



QY 309 ---RRPAQRVTLDAAFEPQNEKGDAPALRVAAESAAA 345  
Db 308 LGHREPMTRVHSSKSVIQE-----EAFEPHVQSTAAPV 340

RESULT 2  
A45070  
nitrilase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A45070  
R;Kobayashi, M.; Komeda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
J. Biol. Chem. 267, 20746-20751, 1992  
A/Title: Nitrilase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the  
A/Reference number: A45070; MUID:93015976; PMID:1400390  
A/Contents: J1  
A/Accession: A45070  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-366 <KOB>  
A/Cross-references: UNIPROT:Q03217; UNIPARC:UPI000016FF36; GB:D11425; NID:g216933; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:116039, NCBIP:116040)  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 46.1%; Score 832.5; DB 2; Length 366;  
Best Local Similarity 52.2%; Pred. No. 6, 2e-63;  
Matches 166; Conservative 43; Mismatches 108; Indels 1; Gaps 1;

QY 1 MSEPMTKYRGAAVQAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIML 60  
Db 1 MVEYTNTEFKVAAVQAPVWFDAKTVDKTVSIIAEARNGCCELVAFPYVFIIPGYPHIW 60  
QY 61 GAPAWGM-RFVQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAGSLYMGQAI 119  
Db 61 DSPLAGMAKFAVRHYENSLTMDSPHVQRLDPAARDHNIAVVGISERDGSPLYMTQLVID 120  
QY 120 PDGDLIAARRKLPKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAM 179  
Db 121 ADGQLVARRRKLPKPTHYERSVYGEENGSDISYDMPFARLGALNCWEHFPQTLTKYAMYSM 180  
QY 180 DEQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVD 239  
Db 181 HEQVHVASWPSGMSLYQPEVPAFGVDAQLTATRMVALLEGQTFVCTTQVVTPEAHEFFCDN 240  
QY 240 PDKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHYSR 299  
Db 241 DEQRKLIGRGGFARITGPDGRDLATPLAEDEEGILYADIDLSAITLAKQAADPVGHYSR 300  
QY 300 PDVTRLLLDLRPAQRVVT 317  
Db 301 PDVLSLNFNQRRHTTPVNT 318

RESULT 3  
A43470  
aliphatic nitrilase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A43470  
R;Kobayashi, M.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
Biochemistry 31, 9000-9007, 1992  
A/Title: Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase  
A/Reference number: A43470; MUID:93003039; PMID:1390687  
A/Contents: K22  
A/Accession: A43470  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-383 <KOB>  
A/Cross-references: UNIPROT:Q02068; UNIPARC:UPI0000130516; GB:D12583; NID:g216931; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:114184, NCBIP:114185)  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 45.7%; Score 825.5; DB 2; Length 383;  
Best Local Similarity 47.2%; Pred. No. 2, 6e-62;  
Matches 169; Conservative 49; Mismatches 125; Indels 15; Gaps 6;  
QY 2 SEPMTKYRG---AAVQAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFW 57  
Db 3 SNPELKYTGKVKVATVQAEPIVLDADATIDKAI GFIEBAAKNGAEFLAFPEVWIPGYPYW 62  
QY 58 IWLGA PWGMR-FVQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAGSLYMGQA 116  
Db 63 AWIGDVKVAVSDFIPKYHENSLTLGDDRMRLQLAARQNNIALVMGYSEKDGASRYLSQV 122  
QY 117 IFGPDGDLIAARRKLPKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAM 176  
Db 123 FIDQNGDIVANRRKLPKPTHYERTYGEENGDTFLTHDFGFRVGGINCWHEHFOPLSKYMM 182  
QY 177 YAADEQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVL 236  
Db 183 YSLNEQIHVASWPAMFALTPDVHQLSVEANDVTVRSYAIEGQTFVLAETHVIGKATQDLF 242  
QY 237 VDTPD-KEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAG 295  
Db 243 AGDDDAKRALPLPGQGWARTYGPDGSLAEPLPEDAEGLYAELDLQIILAKAADPAG 302  
QY 296 HYSRPDVTRLLDRL---PAQRVTLDAAFEPQNEKGDAPALRVV-----ESAAAA 345  
Db 303 HYSRPDVLSLKIDTRNHTPVQ-YITADGRITSLNSNSRVENYRLHQLADIEKENAEAA 359

RESULT 4  
A28658  
nitrilase (EC 3.5.5.1) - Klebsiella ozaenae  
C/Species: Klebsiella ozaenae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C/Accession: A28658  
R;Stalker, D.M.; Malyj, L.D.; McBride, K.E.  
J. Biol. Chem. 263, 6310-6314, 1988  
A/Title: Purification and properties of a nitrilase specific for the herbicide bromoxynil  
A/Reference number: A28658; MUID:88198177; PMID:2834373  
A/Accession: A28658  
A/Molecule type: DNA  
A/Residues: 1-349 <STA>  
A/Cross-references: UNIPROT:P10045; UNIPARC:UPI000013051A; GB:J03196; NID:g149174; PIDN:  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Query Match 34.1%; Score 616; DB 1; Length 349;  
Best Local Similarity 41.1%; Pred. No. 1, 4e-44;  
Matches 139; Conservative 64; Mismatches 113; Indels 22; Gaps 9;

QY 6 TKYRGAAVQAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPAW 65  
Db 3 TTFKAAAVQAEPYWMDAATADKTVTLVAKAAAGQLVAFPELWIPGYGFMULTHNQTE 62  
QY 66 GMRFVQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAGSLYMGQALFGPDGLI 125  
Db 63 TLPFIITKYRKQALAADGPEIEIKIRCAQOEHNIALSFGYSERAGRTLVMGMLIDADGITK 122  
QY 126 AARRKLPKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMVADQVHV 185  
Db 123 IRRRLKPTRFERELFEEDGSDLVQAQTSVGRVYALNCAENLQSLNKFALAAEGEQIHI 182  
QY 186 ASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKV-LVDTPDKE 243  
Db 183 SAWP-FTL--GSPVLVDSIG-AINQVYAETGTFVLMSTQVVGPTGIAAFRIEDRYNPN 238  
QY 244 MFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHYSRPDVT 303  
Db 239 QYL--GGGYARITGPDMLKSKLSPTEEGIYVAEIDLMLAAKYSLDPTGHYSRPDV 296  
QY 304 RLILLD--RRPAQRVTL-----AAFEQNEKGD 331  
Db 297 SVSINRQRPVAVSEV-IDENGDEDPRACEP---DEGD 330

```

RESULT 5
JC4212
nitrilase (EC 3.5.5.1) - Comamonas testosteroni
C/Species: Comamonas testosteroni
C/Date: 14-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 31-Dec-2004
C/Accession: J04212; PC4056
R/Levy-Schli, S.; Soubrier, F.; Crutzz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D.
Gene 161, 15-20, 1995
A/Title: Aliphatic nitrilase from a soil-isolated Comamonas testosteroni sp.: Gene cloning
A/Reference number: J04212; MUID:95369726; PMID:7642130
A/Accession: J04212
A/Molecule type: DNA
A/Residues: 1-354 <LEV>
A/Cross-references: UNIPROT:Q59329; UNIPARC:UPI00000B98BB; GB:L32589; NID:g1082008; PIDN
A/Accession: PC4056
A/Molecule type: protein
A/Residues: 154-162;295-302;323-340 <LE2>
A/Cross-references: UNIPARC:UPI0000175E32; UNIPARC:UPI0000175E33; UNIPARC:UPI0000175E34
C/Comment: This enzyme is active on adiponitrile and cyanovaletric acid.
C/Genetics:
A/Gene: nltA
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase
F;163/Active site: Cys #status predicted

Query Match          32.8%; Score 592.5; DB 2; Length 354;
Best Local Similarity 39.4%; Pred. No. 1.4e-42;
Matches 137; Conservative 51; Mismatches 133; Indels 27; Gaps 7;

QY 11 AAVQAAPVFLDLDRIVEKAIGLIBQAKQDVRLIAPPETWIPGYPFWIWG-----APAW 65
    |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 10 AAVQAAPVFMNLEATVDTCKLIAEASMGAKVIGFPEAFIPGYPIWITSNMDFTGMMW 69
    |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 66 GMRFVQRYFFENSLVRGSKQWQALDAARRHGMHVAVAGYSERAGGSLYMGQATFGPDGLI 125
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 70 AV-----LFKNAIEIPSKVEVQISDAKKNGVYVCYSSEKDNASLYLTQLWFPDNGNLI 124
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 126 AARRKLKPTHAERTVGEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQVHV 185
    |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 125 GKHKRFKPTSSERAIVGDDGSMAPVFKTEYGNLGLQCWEHALPINIAAMGSLNEQVHV 184
    |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 186 ASWPSF-----SLYRGMAVYALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVD 238
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 185 ASWPAFVFKGAVSSRVSSVCASTNAMHQIISQFYAISNQVYVIMSTNLVGDMDIMIGK 244
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 239 TPDKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADPAGHYS 298
    :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 245 DEFSKNFLPLGSGNTAIIISNTGEILAS-IPQDAEGIAVAEIDLNQIITYGKWLIDPAGHYS 303
    :|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 299 RPDVTRLRLDDR--PAQRV---VTLDAAFEPOEDKGD--APALRV 337
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 304 TPGFLSLTFDQSEHVPVKKIGEQTNHFTSYEDLHEDKMDMLTIPRRV 351
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 6
S77025
nitrilase (EC 3.5.5.1) - Synechocystis sp. (strain PCC 6803)
N/Alternate names: protein s110784
C/Species: Synechocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 31-Dec-2004
C/Accession: S77025
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S77025
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA

```

A;Residues: 1-346 <KAN>  
 A;Cross-references: UNIPROT:Q55949; UNIPARC:UPI00000D7133; EMBL:D64005; GB:AB001339; NID  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Genetics:  
 A;Start codon: GTG  
 C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
 C;Keywords: hydrolase

Query Match 32.4%; Score 584.5; DB 2; Length 346;  
 Best Local Similarity 42.1%; Pred. No. 6.5e-42;  
 Matches 128; Conservative 53; Mismatches 114; Indels 9; Gaps 6;

QY 9 RGAAVQAAPVFLDLDRTEVKAIGLIEQAAKQDVRLLAFPETWIPGYPFWIMLGAPA-WGM 67  
 14 RAAAQISFVLFSQGGTMEKVLDAIANAAKGVLLVFPETFPYYPYFSFVEPPVLMGK 73  
 QY 68 RFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGSLYMGQALFGPDGLIAA 127  
 74 SHLKLTYQEAVTVPG-KVTQALIAQAAKTHGMVVVLGVNREREGSLYNTQLIFDADGALVLK 132  
 QY 128 RRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAADQVHVAS 187  
 133 RRKITPTYHERMWGQGDGAGLRTVDTTVGRLGALACWEHYNPLARYALMAQHEQIHCGQ 192  
 QY 188 WPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDPDKEMFLK 247  
 193 FPGSMV--GQIFADQMEV---TWRHHALESQCFVINATGWLTAEQ-KLQITTDEKMHQAL 246  
 QY 248 AGGGFAMIFGPDGRALAEPLPEEGLLVADIDLGMIALAKAADBPAGHSRDPVTRL 307  
 247 SGGCYTAIISPEGKHLCEPIAE-GEGLAIADLDLSLIAKRKRMDSVGHYARPDLLQLTL 305  
 QY 308 DRRP 311  
 DB 306 NNQP 309

RESULT 7  
 JQ1613  
 cyanide hydratase (EC 4.2.1.66) - imperfect fungus (Gloeocercospora sorghi)  
 C;Species: Gloeocercospora sorghi  
 C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004  
 C;Accession: JQ1613  
 R;Wang, P.; VanEtten, H.D.  
 Biochem. Biophys. Res. Commun. 187, 1048-1054, 1992  
 A;Title: Cloning and properties of a cyanide hydratase gene from the phytopathogenic fun  
 A;Reference number: JQ1613; MUID:92412068; PMID:1382413  
 A;Accession: JQ1613  
 A;Molecule type: DNA  
 A;Residues: 1-368 <WAN>  
 A;Cross-references: UNIPROT:P32964; UNIPARC:UPI0000128C2E; GB:M99044; NID:g168196; PIDN  
 C;Comment: This enzyme converts HCN to formamide.  
 C;Genetics:  
 A;Gene: Cht  
 C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
 C;Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase  
 F;286/Binding site: carbonyhydrate (Asn) (covalent) #status predicted

Query Match 31.1%; Score 562; DB 2; Length 368;  
 Best Local Similarity 40.7%; Pred. No. 5.7e-40;  
 Matches 132; Conservative 50; Mismatches 124; Indels 18; Gaps 7;

QY 4 PMTKYRGAAVQAAPVFLDLDRTEVKAIGLIEQAAKQDVRLLAFPETWIPGYPFWIMLGAP 63  
 DB 2 PINKYKAAVVTSEPWENLEGVWKTIEFINEAGKAGCKLIAFPEWIPGYPYMMKVNY 61  
 QY 64 AWGMRFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGSLYMGQALFGPDGD 123  
 DB 62 LQSLPMLKATYRENSIAMDSEMRRIARAARDNQIVSISVSEIDHATLYLTQVLISPLGD 121  
 QY 124 LIAARRKLKPTHAERTVFEGDGS-SHLAVHDTAIGRLGALCWEHIQPLSKYAMYAADQ 182  
 DB 122 VINHRRKIKPTHVEKLVYGDGSDSPFPVTQTEIGRLGQLCWENMNPFLKSLAVARGEQ 181



QY 183 VHVASWPSFSLYRGMAVALGPEVNT----AASQI-----YAVEGGCYVLASCATVSPENIK 234  
Db 182 IHVAAMP---VYPDLRSKVHPDPATNYADPASDLVTPAYAIETGTWVLAFFQRI SVEGLK 238  
QY 235 VLVDTP---DKEMFLKAGGFFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAA 291  
Db 239 --RHTPPGVEPETDATPYNGHARIFRPDGLSLYAKPAVDF-DGLMYVDIDIDNESHLTKALA 295  
QY 292 DPAGHYSRPDVTRLILDRRPAQRV 315  
Db 296 DFAGHYMRPDLIRLLVDTIRKELV 319

RESULT 8

T49148  
nitrlase (EC 3.5.5.1) 3 [imported] - Arabidopsis thaliana  
N/Alternate names: protein T10D17.110  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text\_change 31-Dec-2004  
C/Accession: T49148; T52261; T52264  
R/D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L  
submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z25017  
A/Accession: T49148  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-346 <DAN>  
A/Cross-references: UNIPROT:P46010; UNIPARC:UPI000000BDB6; EMBL:ALJ53865; GSPDB:GN00061;  
A/Experimental source: cultivar Columbia; BAC clone T10D17  
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A/Reference number: Z26007; MUID:98145459; PMID:9484465  
A/Accession: T52261  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-346 <HIL>  
A/Cross-references: UNIPARC:UPI000000BDB6; EMBL:X07648; PIDN:CAA68936.2  
A/Experimental source: cultivar Columbia  
R/Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A/Title: Differential regulation of an auxin-producing nitrlase gene family in Arabidop  
A/Reference number: Z24515; MUID:8022831; PMID:8022831  
A/Accession: T52264  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-346 <BAR>  
A/Cross-references: UNIPARC:UPI000000BDB6; EMBL:U09959; PIDN:AAA19627.1  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: ATSP:T10D17.110; nit3  
A/Map position: 3  
A/Introns: 44/1; 104/1; 202/1; 296/1  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Query Match 25.9%; Score 468.5; DB 2; Length 346;  
Best Local Similarity 33.0%; Pred. No. 4.5e-32;  
Matches 115; Conservative 65; Mismatches 121; Indels 47; Gaps 9;

QY 3 EPMTKYRGAAVQAPVFLDLDRIVEKAIGLEQAQKQDVRLIAFPETWIPGYPFWIMLGA 62  
Db 20 DPSTVTRVTIVQSSVYNDTPATLDKAEKFIWEAASKGAKLVLFPEAFITGYF----- 72  
QY 63 PAWGMRF-----VQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAG 108  
Db 73 --RGFRFGLAVGVHNEGRDEFKRYHASAIKVPGEVERLAELAGKNNVHLYMGAIEKDG 130  
QY 109 GSLYMGQAI FGPDDGLIAARRKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHI 168  
Db 131 YTLYCTALFFSPQGQFLGKHKRKVMPTSLERCIMGQGDSTIPVYDTPIGIKIGAAICWENR 190

QY 169 QPLSKYAMTADEQVHVASWPSFSLYRGMAVALGPEVNTAASQIY-AVEGGCYVLAS--- 224  
Db 191 MPLYRTALYAKGIEIYCAPTADYSL-----EQQASMIHIAVEGGCFVLSAHQF 238  
QY 225 CATVS-PE---MIKVLVDPDKEMFLKAGGFFAMIFGPDGRALAEPLPETEEGLLVADI 279  
Db 239 CKRREFPEHPDYLFNDIVDTKEHDPTVSGG--SVIISPLGKVLAGPNYES-EGLVADL 295  
QY 280 DLGMIALAKAAADPAGHYSRPDVTRLILDRRPAQRVTLDAAPFQNE 327  
Db 296 DLGDIARAKLYFDVGVGHYSKPDIFNLTVNEHPKPPVTFTWKEKADE 343

RESULT 9

T49147  
nitrlase (EC 3.5.5.1) 1 [imported] - Arabidopsis thaliana  
N/Alternate names: protein T10D17.100  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text\_change 31-Dec-2004  
C/Accession: T49147; T52260; T52263  
R/D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L  
submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z25017  
A/Accession: T49147  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-346 <DAN>  
A/Cross-references: UNIPROT:P32961; UNIPARC:UPI0000001AE2; EMBL:ALJ53865; GSPDB:GN00061;  
A/Experimental source: cultivar Columbia; BAC clone T10D17  
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A/Reference number: Z26007; MUID:98145459; PMID:9484465  
A/Accession: T52260  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-346 <HIL>  
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAA68935.2  
A/Experimental source: cultivar Columbia  
R/Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A/Title: Differential regulation of an auxin-producing nitrlase gene family in Arabidop  
A/Reference number: Z24515; MUID:8022831; PMID:8022831  
A/Accession: T52263  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-346 <BAR>  
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:U38845; PIDN:AAB05221.1  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: ATSP:T10D17.100; nit1; NIT1  
A/Map position: 3  
A/Introns: 44/1; 104/1; 202/1; 296/1  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Query Match 25.9%; Score 467; DB 2; Length 346;  
Best Local Similarity 34.1%; Pred. No. 6.1e-32;  
Matches 121; Conservative 50; Mismatches 122; Indels 62; Gaps 8;

QY 4 PMTKYRGAAVQAPVFLDLDRIVEKAIGLEQAQKQDVRLIAFPETWIPGYPFWIMLGA 63  
Db 21 PSTTVTRVTIVQSSVYNDTPATIDKAEKFIWEAASKGAELVLFPEGFITGYF----- 72  
QY 64 AWGMRF-----VQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAG 109  
Db 73 --RGFRFGLAVGVHNEGRDEFKRYHASAIHVPGEVARLADYARKNHVYLLVMGAIEKEGY 131  
QY 110 SLYMGQAI FGPDDGLIAARRKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQ 169  
Db 132 TLYCTVLFSPQGQFLGKHKRKLMPISLERCIMGQGDSTIPVYDTPIGIKLGAICWENRM 191  
QY 170 PLSKYAMTADEQVHVA-----SWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVL 222

Db 192 PLVRTALYAKGIELYCAPTADGSKEWQSSMLH-----IAIEGGCFVL 233  
QY 223 ASCATVSPEMIKVLVDTP-----DKEMFLKAGGFAMIFGPDGRALAEPLPETEEG 273  
Db 234 SACQFCQR---KHFPDHPDYLFDTWYDDKEHDSIVSQGGSVIISPLGQVLAGPNFES-EG 289  
QY 274 LLVADIDLGMIALAKAAADBPAGHYSRPDVTRL.LDRRPAQRVVTLDAAFEPOVED 328  
Db 290 LVTADIDLGDIAARAKLYFDSVGHYSRDPVLT.LLTVNEHP-RKSVTFVTKEKAEDD 343

RESULT 10  
S31969  
nitrilase (EC 3.5.5.1) [similarity] - Arabidopsis thaliana  
N;Alternate names: T10D17.90  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C/Accession: S31969; T49146; T52258  
R;Barling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.  
submitted to the EMBL Data Library, September 1992  
A/Description: Arabidopsis thaliana nitrilases are encoded by a two-member gene family:  
A/Reference number: S31969  
A/Accession: S31969  
A/Molecule type: mRNA  
A/Residues: 1-339 <BAR>  
A/Cross-references: UNIPROT:P32962; UNIPARC:UPI0000001444; EMBL:X68305; NID:g22655; PID:  
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I  
submitted to the Protein Sequence Database, April 2000  
A/Reference number: Z25017  
A/Accession: T49146  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-339 <DAN>  
A/Cross-references: UNIPARC:UPI0000001444; EMBL:AL353865; GSPDB:GN00061; ATSP:T10D17.90  
A/Experimental source: cultivar Columbia; BAC clone T10D17  
R;artling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6021-5, 1994  
A/Title: Molecular characterization of two cloned nitrilases from Arabidopsis thaliana:  
A/Reference number: Z24514; MUID:8016109; PMID:8016109  
A/Accession: T52258  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-339 <BA2>  
A/Cross-references: UNIPARC:UPI0000001444; EMBL:X68305; PIDN:CAA48377.1  
A/Experimental source: cultivar Landsberg erecta  
C/Genetics:  
A/Gene: ATSP:T10D17.90  
A/Map position: 3  
A/Introns: 37/1; 97/1; 195/1; 289/1  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Query Match 25.7%; Score 464; DB 2; Length 339;  
Best Local Similarity 34.6%; Pred. No. 1.1e-31;  
Matches 119; Conservative 54; Mismatches 127; Indels 44; Gaps 8;  
QY 6 TKYGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYP--FWIWLGA 63  
Db 16 TIVRATIVQASTVYNDTPATLEKANKFIVEAASKSELVVFPEAFIGYPRGFRFLGVG 75  
QY 64 AW---GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSLYMGQALFGP 120  
Db 76 VHNIEGRDEFKRYHASAIKVPGEVEKLAELAGKNNVYLVMGALIEKDGTYLYCTALFFSP 135  
QY 121 DGDLLIARRKCLKPTHAERTVFEGBDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 136 QGQFLGKHKRLMPTSLERCISWGQDGSTIPVYDTPIGKLGAAICWENRMPLYRTALYAKG 195  
QY 181 EQVHVA-----SWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYTLASCATVSPEMI 233  
Db 196 IELYCAPTADGSKEWQSSMLH-----IAIEGGCFVLSACQFC---LR 234

QY 234 KVLVDTP-----DKEMFLKAGGFAMIFGPDGRALAEPLPETEEGLVADIDLGMI 284  
Db 235 KDFPDHPDYLFDTWYDDKEPDSIVSQGGSVIISPLGQVLAGPNFES-EGLTADLDLGDV 293  
QY 285 ALAKAAADBPAGHYSRPDVTRL.LDRRPAQRVVTLDAAFEPOVED 328  
Db 294 ARAKLYFDSVGHYSRDPVLT.LLTVNEHP-KKPVTFISKVEKAEDD 336

RESULT 11  
T52259  
nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52259  
R;Hillebrand, H.; Barling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A/Reference number: Z26007; MUID:98145459; PMID:9484465  
A/Accession: T52259  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-339 <HIL>  
A/Cross-references: UNIPROT:O04907; UNIPARC:UPI00000A5899; EMBL:X07648; PIDN:CAA68934.3  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: nit2  
A/Introns: 37/1; 97/1; 195/1; 289/1  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Query Match 25.6%; Score 463; DB 2; Length 339;  
Best Local Similarity 34.6%; Pred. No. 1.3e-31;  
Matches 119; Conservative 54; Mismatches 127; Indels 44; Gaps 8;  
QY 6 TKYGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYP--FWIWLGA 63  
Db 16 TIVRATIVQASTVYNDTPATLEKANKFIVEAASKSELVVFPEAFIGYPRGFRFLGVG 75  
QY 64 AW---GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSLYMGQALFGP 120  
Db 76 VHNIEGRDEFKRYHASAIKVPGEVEKLAELAGKNNVYLVMGALIEKDGTYLYCTALFFSP 135  
QY 121 DGDLLIARRKCLKPTHAERTVFEGBDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 136 QGQFLGKHKRLMPTSLERCISWGQDGSTIPVYDTPIGKLGAAICWENRMPLYRTALYAKG 195  
QY 181 EQVHVA-----SWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYTLASCATVSPEMI 233  
Db 196 IELYCAPTADGSKEWQSSMLH-----IAIEGGCFVLSACQFC---LR 234  
QY 234 KVLVDTP-----DKEMFLKAGGFAMIFGPDGRALAEPLPETEEGLVADIDLGMI 284  
Db 235 KDFPDHPDYLFDTWYDDKEPDSIVSQGGSVIISPLGQVLAGPNFES-EGLTADLDLGDV 293  
QY 285 ALAKAAADBPAGHYSRPDVTRL.LDRRPAQRVVTLDAAFEPOVED 328  
Db 294 ARAKLYFDXVGHYSRDPVLT.LLTVNEHP-KKPVTFISKVEKAEDD 336

RESULT 12  
S22398  
nitrilase (EC 3.5.5.1) - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C/Accession: S22398  
R;Barling, D.; Seedorf, M.; Mithoefer, A.; Weiler, E.W.  
Eur. J. Biochem. 205, 417-424, 1992  
A/Title: Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-a  
A/Reference number: S22398; MUID:92209532; PMID:1555601  
A/Accession: S22398  
A/Molecule type: mRNA  
A/Residues: 1-346 <BAR>

A;Cross-references: UNIPROT:P32961; UNIPARC:UPI0000130513; EMBL:X63445; NID:g16399; PIDN  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase

Query Match 25.5%; Score 461; DB 2; Length 346;  
Best Local Similarity 33.8%; Pred. No. 2e-31;  
Matches 120; Conservative 51; Mismatches 122; Indels 62; Gaps 8;

QY 4 PMTKYGAAVQAAVFLDLDRTEVKAIGLIEQAAKQDVRLIAFPETWIPGYPFWILGAP 63  
DB 21 PSTTVRVTIVQSSIVYNDTPATIDKAEKVIVEAASKGAELVLPPEGFIGGYP----- 72  
QY 64 AWGMR-----VQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGG 109  
DB 73 -RGFRFGLAVGVHNEGRDEFKRYHASAIHVPGPEVARLADVARKNHVYLVGALIEKEGY 131  
QY 110 SLVMQAIFFPDGLIAARRKLPHTAERTVFEGDGSHLAVHDTAIGRLGALCCWEHIQ 169  
DB 132 TLVCTVLFSPQGFGLGKRKLMPTSLERCIVGQGDGSTIPVYDTPIGKLGAALCWEENRM 191  
QY 170 PLSKYAMYAADEQVHA-----SWPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVL 222  
DB 192 PLYRTALYAKGIELYCAPTADGSKEWQSSMLH-----IAIEGGCFVL 233  
QY 223 ASCATVSPMIKVLVDP-----DKEMFLKAGGFAMIFGPDGRALAEPLPETE 273  
DB 234 SACQFCQR--KHPDPHDYLFDTWYDDKEHDSIVSQGSVITISPLQVLAGEPNFES-EG 289  
QY 274 LLVADIDLGMIALAKAADPAGHYSRPDVTRLIDRRPAQRVVTLDAAFEPO 328  
DB 290 LVTADIDLGDIAAKLYFDSVGYSRPDVHLTVNEHP-RKSVTFVTKEKAEDD 343

RESULT 13

nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C;Accession: T52262  
R;Barrel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A;Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A;Reference number: Z24515; MUID:8022831; PMID:8022831  
A;Accession: T52262  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-339 <BAR>  
A;Cross-references: UNIPROT:P32962; UNIPARC:UPI000016DA4C; EMBL:U38845; PIDN:AAB05220.1  
C;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: NIT2  
A;Map position: 3  
A;Intons: 37/2; 97/1; 195/1; 289/1  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase

Query Match 25.2%; Score 456; DB 2; Length 339;  
Best Local Similarity 34.3%; Pred. No. 5.1e-31;  
Matches 118; Conservative 54; Mismatches 128; Indels 44; Gaps 8;

QY 6 TKYGAAVQAAVFLDLDRTEVKAIGLIEQAAKQDVRLIAFPETWIPGYP--FWIWLGA 63  
DB 16 TIVRATIVQASTVYNDTPATLKGANKFIVEAATKSELVVFPEAFIGGYPRFRFGLGV 75  
QY 64 AW---GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLVMQAI 120  
DB 76 VHNNEGRDEFKRYHASAIKVPGPEVEKLAELAGKNVYLVGALIEKDGTYLYCTALFFSP 135  
QY 121 DGDLLIARRKLPHTAERTVFEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180  
DB 136 QGQFLGKRKLMPTSLERCIVGQGDGSTIPVYDTPIGKLGAALCWEENRMPLYRTALYAKG 195  
QY 181 EQVHVA-----SWPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSP 233

DB 196 IELYCAPTADGSKEWQSSMLH-----IAIEGGCFVLSACQFC--LR 234  
QY 234 KVLVDTP-----DKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGM 284  
DB 235 KDFPDHPDYLFDTWYDDKEPDSIVSQGSVITISPLQVLAGEPNFES-EGLITADLDLGDV 293  
QY 285 ALAKAAADPAGHYSRPDVTRLIDRRPAQRVVTLDAAFEPO 328  
DB 294 ARAKLYFDSVGHYSRPDVHLTVNEHP-KKPVTFISKVEKAEDD 336

RESULT 14

T52266  
nitrilase-like protein [imported] - rice  
C;Species: Oryza sativa (rice)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C;Accession: T52266  
R;Chiba, R.; Dohmoto, M.; Yamaguchi, K.  
submitted to the EMBL Data Library, May 1999  
A;Description: Oryza sativa, a gene for nitrilase-like protein.  
A;Reference number: Z26008  
A;Accession: T52266  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-362 <CHI>  
A;Cross-references: UNIPROT:Q9SXX6; UNIPARC:UPI000009C92A; EMBL:AB027054; PIDN:BA077679.  
A;Experimental source: cultivar Nipponbare  
C;Genetics:  
A;Gene: ONIT4  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 23.3%; Score 420; DB 2; Length 362;  
Best Local Similarity 34.1%; Pred. No. 6.3e-28;  
Matches 117; Conservative 57; Mismatches 137; Indels 32; Gaps 10;

QY 9 RGAAVQAAVFLDLDRTEVKAIGLIEQAAKQDVRLIAFPETWIPGYPFW-----IWL 60  
DB 32 RATVQASTVYFDTPATLIDKAERLIEAAGYSQLVVFPEAFVGYPRGSTFGGANISI 91  
QY 61 GAPA-WGMRFYQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLVMQAI 119  
DB 92 GNPDKKGKEFRKRYHAAALIEVPGPVYTRLAAMAGKYKFLVMGVIEREGYTLVCSV 151  
QY 120 PDGDLIARRKLPHTAERTVFEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAA 179  
DB 152 PLGRYLGKRKLMPTALERIIVGFQDGTIPVYDTPLGKIGALICWENKMPRLRTALYGK 211  
QY 180 DEOVHVASWPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLAS---CATVS----PEM 232  
DB 212 GIEIYCA--PT-----ADSRQWQASMTHTIALEGCFVLISANQFCRRKDYPPPEY 260  
QY 233 IKV-LVDTPRKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAA 291  
DB 261 VFTGLGEPSPDVTVCPPG--SVIISPSGEVLAGEPNYE-GEALITADLDLGEIVRAK 317  
QY 292 DPAGHYSRPDVTRLIDRRPAQRVVTLDAAFEPO 334  
DB 318 DVVGHYARPEVLSLVVNDQ-HLPVSFISAETTAAKSDSTA 359

RESULT 15

T03739  
nitrilase (EC 3.5.5.1) 4B - common tobacco  
C;Species: Nicotiana tabacum (common tobacco)  
C;Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C;Accession: T03739  
R;Tsunoda, H.  
submitted to the EMBL Data Library, January 1996  
A;Reference number: Z15044  
A;Accession: T03739  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:35 ; Search time 155.523 Seconds  
(without alignments)  
1569.628 Million cell updates/sec

Title: US-09-751-299-2  
Perfect score: 1806  
Sequence: 1 MSEPMTKYRGAAGVQAPVFL.....EDKGDAPALRVVAESAASAAQ 346

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	1806	100.0	346	2 Q6RWK5_9ZZZZ	Q6rwk5 uncultured
2	1464	81.1	341	2 Q6RWR4_9ZZZZ	Q6rwr4 uncultured
3	1209	66.9	348	2 Q6RWS2_9ZZZZ	Q6rws2 uncultured
4	1165	64.5	353	2 Q6RWF9_9ZZZZ	Q6rwf9 uncultured
5	1155	64.0	353	2 Q6RWF0_9ZZZZ	Q6rwf0 uncultured
6	1132.5	62.7	358	2 Q6RWF0_9ZZZZ	Q6rwi0 uncultured
7	1130	62.6	354	2 Q6RWK6_9ZZZZ	Q6rwk6 uncultured
8	1127	62.4	331	2 Q706Q8_PSEPU	Q706q8 pseudomonas
9	1125	62.3	345	2 Q6RWS5_9ZZZZ	Q6rws5 uncultured
10	1099.5	60.9	334	2 Q6RWE3_9ZZZZ	Q6rwe3 uncultured
11	1076.5	59.6	337	2 Q6RWN9_9ZZZZ	Q6rwn9 uncultured
12	1073.5	59.4	338	2 Q6RWG6_9ZZZZ	Q6rwg6 uncultured
13	1070	59.2	336	2 Q6RWF7_9ZZZZ	Q6rwf7 uncultured
14	1068.5	59.2	335	2 Q5YUM5_NOCFA	Q5yum5 nocardia fa
15	1066.5	59.1	337	2 Q6RWN8_9ZZZZ	Q6rwn8 uncultured
17	1064.5	58.9	338	2 Q6RWN7_9ZZZZ	Q6rwn7 uncultured
18	1064.5	58.9	350	2 Q5EG61_PSEFL	Q5eg61 pseudomonas
19	1057.5	58.6	352	2 Q6RWH0_9ZZZZ	Q6rwh0 uncultured
20	1052.5	58.3	338	2 Q6RWS4_9ZZZZ	Q6rws4 uncultured
21	1044.5	57.8	337	2 Q6RWS3_9ZZZZ	Q6rws3 uncultured
22	1043.5	57.8	338	2 Q6RWM2_9ZZZZ	Q6rwm2 uncultured
23	1041	57.6	336	2 Q6RWM5_9ZZZZ	Q6rwm5 uncultured
24	1038.5	57.5	338	2 Q6RWQ4_9ZZZZ	Q6rww4 uncultured
25	1038.5	57.5	338	2 Q6RWR6_9ZZZZ	Q6rwr6 uncultured
26	1028.5	56.9	337	2 Q6RWI7_9ZZZZ	Q6rwi7 uncultured
27	1026.5	56.8	338	2 Q6RWH1_9ZZZZ	Q6rwh1 uncultured
28	1025.5	56.8	336	2 Q500U1_PSESY	Q500u1 pseudomonas
29	1023	56.6	337	2 Q6RWQ3_9ZZZZ	Q6rwq3 uncultured
30	1023	56.6	337	2 Q6RWQ7_9ZZZZ	Q6rwq7 uncultured
31	1020.5	56.5	327	2 Q6RWH9_9ZZZZ	Q6rwh9 uncultured

32	1012	56.0	347	2 Q88B32_PSESM	Q88b32 pseudomonas
33	1008	55.8	349	2 Q6RWS0_9ZZZZ	Q6rws0 uncultured
34	1007	55.8	337	2 Q6RWE0_9ZZZZ	Q6rwe0 uncultured
35	1004.5	55.6	337	2 Q6RWM3_9ZZZZ	Q6rwm3 uncultured
36	986	54.6	345	2 Q6RWN2_9ZZZZ	Q6rwn2 uncultured
37	974.5	54.0	337	2 Q6RWM9_9ZZZZ	Q6rwm9 uncultured
38	972	53.8	334	2 Q6RWG9_9ZZZZ	Q6rwg9 uncultured
39	964	53.4	334	2 Q89GE3_BRAJA	Q89ge3 bradyrhizob
40	949.5	52.6	346	2 Q6RWM1_9ZZZZ	Q6rwm1 uncultured
41	949	52.5	333	2 Q6RWL0_9ZZZZ	Q6rwl0 uncultured
42	945	52.3	332	2 Q6RWJ0_9ZZZZ	Q6rwj0 uncultured
43	869	48.1	356	1 NRLA_ALCFA	P20960 alcaligenes
44	829.5	45.9	365	1 NRL2_RHORH	Q03217 rhodococcus
45	825.5	45.7	383	1 NRL1_RHORH	Q02068 rhodococcus

ALIGNMENTS

RESULT 1					
ID	Q6RWK5_9ZZZZ	PRELIMINARY;	PRT;	346 AA.	
AC	Q6RWK5;				
DT	05-JUL-2004 (Tremblrel. 27, Created)				
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)				
DT	05-JUL-2004 (Tremblrel. 27, Last annotation update)				
DE	Nitrilase (EC 3.5.5.7).				
GN	ORFNames=BD5086;				
OS	uncultured organism.				
OC	unclassified; environmental samples.				
OX	NCBI_TaxID=155900;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RX	PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;				
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,				
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,				
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,				
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.;				
RT	"Exploring nitrilase sequence space for enantioselective catalysis.";				
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).				
DR	EMBL; AY487497; AAR97444.1; -; Genomic_DNA.				
DR	GO; GO:0018762; F:aliphatic nitrilase activity; IEA.				
DR	GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.				
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.				
DR	InterPro; IPR00132; N/CN_hydrtase.				
DR	InterPro; IPR003010; Ntlase/CNhydrtse.				
DR	Pfam; PF00795; CN_hydrolyase; 1.				
DR	PROSITE; PS50263; CN_HYDROLASE; 1.				
DR	PROSITE; PS00921; NITRIL_CHT_2; 1.				
KW	Hydrolyase.				
SQ	SEQUENCE 346 AA; 37432 MW; 12FE01C35AD463CE CRC64;				
Query Match 100.0%; Score 1806; DB 2; Length 346;					
Best Local Similarity 100.0%; Pred. No. 4.5e-130;					
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1 MSEPMTKYRGAAGVQAPVFLDLDRTEVEKAIGLIEQAQKQDVRLLAFPETWIPGYFWIWL 60				
DB	1 MSEPMTKYRGAAGVQAPVFLDLDRTEVEKAIGLIEQAQKQDVRLLAFPETWIPGYFWIWL 60				
QY	61 GAPAWGMRFPVQRYFENSLVRGSKQWQALADARRHGHWVAVGYSERAGGSLVMGQALFGP 120				
DB	61 GAPAWGMRFPVQRYFENSLVRGSKQWQALADARRHGHWVAVGYSERAGGSLVMGQALFGP 120				
QY	121 DGDLLIARRKLLKPTAERTVFEEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180				
DB	121 DGDLLIARRKLLKPTAERTVFEEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180				
QY	181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLYVDT 240				
DB	181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLYVDT 240				
QY	241 DKEMFLKAGGFAFIQPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAHGYSRP 300				

Db 241 DKEMFLKAGGPFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300  
QY 301 DVTRLILLDRRPAQRVVTLLDAAFEPQNEDEKGDAPALRVVAESAAAAQ 346  
Db 301 DVTRLILLDRRPAQRVVTLLDAAFEPQNEDEKGDAPALRVVAESAAAAQ 346

RESULT 2  
Q6RWR4\_9ZZZZZ  
ID Q6RWR4\_9ZZZZZ PRELIMINARY; PRT; 341 AA.  
AC Q6RWR4;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7266;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487438; AAR97385.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; NtIse/CNhydrtse.  
DR Pfam; PF00795; CN\_hydrolyase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KM Hydrolyase.  
SQ SEQUENCE 341 AA; 36918 MW; 23AC5387F4C1FDEE CRC64;

Query Match 81.1%; Score 1464; DB 2; Length 341;  
Best Local Similarity 81.7%; Pred. No. 7e-104;  
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;  
QY 1 MSEPTKYRGAAYVQAAVFLDLDRTEVEKAIGLIEQAAKODVRLIAFPETWIPGYPFWIML 60  
Db 1 MLSPVTQYRAAAVQAAVFLDLDRTEVEKTAIEQAAEQDVRLIAFPETWIPGYPLWIML 60  
QY 61 GAPAWGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGP 120  
Db 61 GSPAWGMRVQRYFENSLVRGSKQWNAIADAAARRHMTVVVGFSERAGGSLYMGQAIFGP 120  
QY 121 DGDLLIARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180  
Db 121 EGELIARRKCLKPTHAERTVFGEGDGSHLAVYETGVGRIGALCCWEHIQPLSKYAMYAAN 180  
QY 181 EQYHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTP 240  
Db 181 EQYHVASWPCFSLYRGMAVALGPEVNTAASQYVAVEGGCYVLASCLVVTPEILKVLIDTP 240  
QY 241 DKEMFLKAGGPFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300  
Db 241 DKEPULLLAGGFSMTFGPDGRALAOPLPETEEGLVTAIDLGAIALAKAAADPAGHYARP 300  
QY 301 DVTRLILLDRRPAQRVVTLLDAAFEPQNEDEKGDAP 333  
Db 301 DVTRLILLNPRPAARVEALGPRFEVVOSEQAEPP 333

RESULT 3  
Q6RWS2\_9ZZZZZ  
ID Q6RWS2\_9ZZZZZ PRELIMINARY; PRT; 348 AA.

AC Q6RWS2;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5338;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.

RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487430; AAR97377.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; NtIse/CNhydrtse.  
DR InterPro; IPR002332; PII\_GlnB\_UMP\_S.  
DR Pfam; PF00795; CN\_hydrolyase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN\_1.  
KM Hydrolyase.  
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;

Query Match 66.9%; Score 1209; DB 2; Length 348;  
Best Local Similarity 67.1%; Pred. No. 2.5e-84;  
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;  
QY 4 PMTK-YRGAAYVQAAVFLDLDRTEVEKAIGLIEQAAKODVRLIAFPETWIPGYPFWIMLGA 62  
Db 2 PTSKQFRVAAYVQAAVFLDLLEGALISKGISLIEBAASNGAKLIAFPETWIPGYPFWIMLDS 61  
QY 63 PAMGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGPDG 122  
Db 62 PAMGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQSIINDKG 121  
QY 123 DLIAARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQ 182  
Db 122 ETITRRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQ 181  
QY 183 VHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTPDK 242  
Db 182 IHIASWPSFSLYRGMAVALGPELNNAAEQYAAEGQCFVLAPCATVSKEMIEMLIDDPK 241  
QY 243 EMFLKAGGPFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDV 302  
Db 242 EPLLLEGGFTWITGPDGRPLAKPLPENEEGLLYADIDLGMIAMAKAAADPAGHYARPDV 301  
QY 303 TRLLIDRRRPAQRVVTLLDAAFEPQNEDEKGDAPALRVVAESAAAA 345  
Db 302 TRLLFNSAPANRVEYINPASGP-TESLDMGKMQMEAEQOKAA 343

RESULT 4  
Q6RWF9\_9ZZZZZ  
ID Q6RWF9\_9ZZZZZ PRELIMINARY; PRT; 353 AA.  
AC Q6RWF9;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7753;  
OS uncultured organism.  
OC unclassified; environmental samples.

OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487543; AAR97490.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydrtse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC31326 CRC64;

Query Match 64.5%; Score 1165; DB 2; Length 353;  
Best Local Similarity 69.4%; Pred. No. 6e-81;  
Matches 229; Conservative 31; Mismatches 66; Indels 4; Gaps 3;

QY 7 KYRGAAVQAAPVFLDRLTVEKAIGLIEQAQKQDVRLIAPPETWIPGYFWIWLGAAPWG 66  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 7 KYKVAAVQAAPAFDLIDASVEKAVRFIDEAGAAGARLLIAPPETWIPGYFWIWLGAAPWA 66  
QY 67 -MR-FVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQATFGPDGL 124  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 67 IMRGVSRYPFDSLSDSPQAEKRLAARKNMNVVLGLSERDGSGLYIAQWITIGPDGET 126  
QY 125 IAAARKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQVH 184  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 127 IAKRRKLKPTHAERTVFEEDGSHLAVHDLVGRIGALCCWEHLQPLSKYAMYAQNEQVH 186  
QY 185 VASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPEKEM 244  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 187 VAAMPSSFSLYDPFAHALGAEVNNAASKIYAVEGSCFVIAPCATVSGAMIDELCTPEKHQ 246  
QY 245 FLKAGGCFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAADPAGHYSRPDVTR 304  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 247 FLHAGGFAVITYGPDGAPLAPLPDPKEGLLYADIDLGMISVAKAADPAGHYARPDVTR 306  
QY 305 LLLDRRPAQRVVTLDAAFEPPQNEDEKGDAPA 334  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 307 LLFNNRPGYRVETM--ALPIDAETKAEAPA 334

RESULT 5  
Q6RWF0\_9ZZZZZ PRELIMINARY; PRT; 353 AA.  
ID Q6RWF0\_9ZZZZZ PRELIMINARY; PRT; 353 AA.  
AC Q6RWF0;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).

DR EMBL; AY487552; AAR97499.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydratase.  
DR InterPro; IPR003010; Ntlse/CNhydrtse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 353 AA; 38321 MW; 2AA05257B3A7BC6C CRC64;

Query Match 64.0%; Score 1155; DB 2; Length 353;  
Best Local Similarity 66.9%; Pred. No. 3.5e-80;  
Matches 224; Conservative 38; Mismatches 69; Indels 4; Gaps 3;

QY 1 MSEPMTKYRGAAVQAAPVFLDRLTVEKAIGLIEQAQKQDVRLIAPPETWIPGYFWIWL 60  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 1 MSQIHPKLKVAAVQAAPAFDLIDASIEKTIKYVDEAAAGAKLLIAPPETWIPGYFWIWL 60  
QY 61 GAPAWG-MR-FVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQATF 118  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 61 GAPAWAIMRGFVSRYPFDSLSDSPQAEKRLRDAARRNKIYIALGLSERDGSGLYIAQWII 120  
QY 119 GPDGDLIAARRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYA 178  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 121 GPGGETVAQRRLKPTHAERTVFEEDGSHLAVHDLIDIGRLGALCCWEHLQPLSKYAMYA 180  
QY 179 ADEQVHAVSWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVD 238  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 181 QNEQIHVAAMPSSFSLYDPFAHALGAEVNNAASKIYAVEGSCFVIAPCATVSGVMIDELCD 240  
QY 239 TPDKEMFLKAGGCFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAADPAGHY 298  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 241 TPEKHQFLHVGGFAYIYGPDGSPPLAKPLPPDQEGLLYADIDLGMISVAKAADPAGHYA 300  
QY 299 RPDVTRLDDRPAQRVVTLDAAFEPPQNEDEKGDAP 333  
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 301 RPDVTRLFNNRPAARVEKL--ALPVDQEAEDVSP 333

RESULT 6  
Q6RWIO\_9ZZZZZ PRELIMINARY; PRT; 358 AA.  
ID Q6RWIO\_9ZZZZZ PRELIMINARY; PRT; 358 AA.  
AC Q6RWIO;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5263;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487522; AAR97469.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydrtse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.



[illegible]

Query Match	Best Local Similarity	Matches	Score	DB 2;	Length	Indels	Gaps
62.6%;	64.4%;	224;	1130;	2;	354;	74;	4
Conservative	Pred. No. 2.9e-78;	42;	Mismatches	74;	Indels	8;	Gaps
7 KYRGAAVQAAPVFLDLDRITVEKATGLEQAAKQDVRLIAFPETWIPGYPFWMILGAPAWG	7 KFKVAAVQAAPAFLLDLDAVEKAIVRLIDEAGAGARLIAFPETFTIPGYPFWMILGAPAWA	66	66	66	66	66	66

QY		67	-MR-FVORYFENSLVRGSKOMALADARRHGMHVAGYSERAGSLYMGAIFGPDGL	124
Dd		67	IMRGFVSRYFDNSLOYGTPEADRRLRAAKRNKMFWALGLSERDGSLEYIAQWIIIGPDGET	126
QY		125	IARRKLKPTHAERTVFEGGDGSHLAVHDTAIGRLGALCCWEHIOPLSKYAMTAADEOVH	184
Dd		127	VATRKCLKPTHAERTVFEGGDGSHLAVHELDIGRVGALCCWEHTLOPLSKYAMYAQNEOVH	186
QY		185	VASWPSFSLYRGMAYALGPVENTAASQIYAVEGGCYVLASCATVSPENIKYLVDTPDKEM	244
Dd		187	IAAMPSSFSLYPDFAHALGAEVNNAASKITYAVEGSCFVIAPCATVSCAMIDELCDSPEKHQ	246
QY		245	FLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMTALAKAAADPAGHSRPDVTR	304
Dd		247	FLHVGGGFVITGPDGAPLAKRIPLAPDOEGLLYADIDLGMISVAKAAADPAGHYARPdvTR	306
QY		305	LILDRPAQRVVTLDAAFEPONE-DKGDAPALRVVAESA-----AAAO	346
Dd		307	LLENRRPGNRVETLALPVDDQEAEGAGAGKPAKPSPSVAAFTLTQAALAE	354

```

RESULT 8
Q706Q8_PSEPU PRELIMINARY; PRT; 331 AA.
ID Q706Q8_PSEPU
AC Q706Q8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Predicted amidohydrolase/nitriase.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RX MEDLINE=99008981; PubMed=9791097;
RA Ravatn R., Studer S., Zender A.J.B., van der Meer J.R.;
RT "Int-B13, an unusual site-specific recombinase of the bacteriophage P4
RT integrase family is responsible for chromosomal insertion of the 105-
RT kb c1c-element of Pseudomonas sp. strain B13.";
RL J. Bacteriol. 180:5505-5514(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RX MEDLINE=22708816; PubMed=12823813;
RX DOI=10.1046/j.1365-2958.2003.03548.x;
RA Sentschilo V., Zehnder A.J.B., van der Meer J.R.;
RT "Characterization of two alternative promoters and a transcription
RT regulator for integrase expression in the clc catabolic genomic island
RT of Pseudomonas sp. strain B13.";
RL Mol. Microbiol. 49:93-104(2003).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=strain RR21;
RA Gaillard M., Werlen C., Vallaes T., Vorhoelter F.J., Puehler A.,
RA van der Meer J.R.;
RA Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ617740; CAE92892.1; -; Genomic_DNA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN_hydrtase.
DR InterPro; IPR003010; Ntise/CNhydrtse.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 331 AA; 36554 MW; DEF59D03874D9187 CRC64;

Query Match 62.4%; Score 1127; DB 2; Length 331;
Best Local Similarity 65.2%; Pred. No. 4.5e-78;

```

Matches	206;	Conservative	45;	Mismatches	65;	Indels	0;	Gaps	0
QY	5	MTKRYGAAVQAAPVFLDLDRIVEKAIGLEQAAKODVRLIAFPETWIPGYPFWIMLGAPA	64						
Db	6	IAKYKVAAVQAAPEFLNLDKGVEKAVRLIEEAAKNGAKLIAFPPEWLPGYPMWIMLDSPA	65						
QY	65	WGMRFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGSLYMGQAIFGPDGDL	124						
Db	66	WGMQFVQRHFENALLVGSPOWERLCAAAAEHRI FVVLGFCERQDGTLYIAQAIIDDEGRV	125						
QY	125	IAARRKLKPTHAERTVFEGEGDGHSLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQVH	184						
Db	126	VSTRKKLKPTHAERTVYGEEDGSHLSVHOTSIGRIGALSCAEHIQPLSKYAMYSQNEQIH	185						
QY	185	VASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPDKEM	244						
Db	186	VAAWPSFSVYRGAAFQLSPKANLTAASQVYALEGGCYVLAPCALVSKEMLEMLADTPQKRQ	245						
QY	245	FLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHYSRPDVTR	304						
Db	246	LLLEGGGFAQIFGPDAKPLCEPFPETQEGLLYADVDLGFIGVAKAAYDPTGHYSRPDVVR	305						
QY	305	LLIDRRPAQRVVTLDA	320						
Db	306	LLMNSKPAITRVHSFOA	321						

```

RESULT 9
O6RMS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
ID O6RMS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
AC O6RMS5;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DE 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DI NitriIase (EC 3.5.5.7).
GN ORFNames=BD7611;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chet E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitriIase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL: AY487427; AAR97374.1; -; Genomic_DNA.
DR GO: GO:0018762; F:aliphatic nitriIase activity; IEA.
DR GO: GO:0016810; F:hydroIase activity, acting on carbon-nitrog. . . ; IEA.
DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro: IPR000132; N/CN hydrtase.
DR InterPro: IPR003010; NIIsE/CNHytIsE.
DR Pfam; PF00795; CN_hydroIase; 1.
DR PROSITE; PS0263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRII_CHT_1; 1.
DR PROSITE; PS00921; NITRII_CHT_2; 1.
KW HydroIase.
SQ SEQUENCE 345 AA; 37242 MW; 5F8A3CFE530D2F1D CRC64;

Query Match 62.3%; Score 1125; DB 2; Length 345;
Best Local Similarity 63.9%; Pred. No. 6.7e-78;
Matches 212; Conservative 44; Mismatches 72; Indels 4; Gaps 11
```

```

QY      127 ARRLKPTHAERTVFEBGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADQOVHVA 186
      ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db      127 TRRLKPTHERAIFEGEGDSDLAVHDTKLGRVGLCCWEHLQPLSKYAMYAONEQVHIA 186
QY      187 SWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLDTPDKEMFL 246
      :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :
Db      187 AWPFSFLYVDAAYALGPEVNNAASRLYAVEGQCFFVAPCATVSOQKIDMLCETPEQOALL 246
QY      247 KAGGGFAMIFGPDGRALAEPLPTEEGLLVADIDLGMTALAKAADPAGHSRPDVRTLL 306
      ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||||| :|||: ||
Db      247 KPGGGAHQIYGPDGRLADPLPDAEGLLYADIDLAAITLAKAADPAGHSRPDVTQLL 306
QY      307 LDRRPAQRVVTLLDAFEPQNEDEKGDAPALRVV 338
      ||| | ||| : | : |||
Db      307 LDRNPKPRVVHA---KPGGSANNSSPGMRAY 334

```

	RESULT	10
ID	Q6RWE3_9ZZZZZ PRELIMINARY;	PRT; 334 AA.
AC	Q6RWE3;	
DT	05-JUL-2004 (TREMBlrel. 27, Created)	
DT	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	
DE	Nitrilase (EC 3.5.5.7).	
OS	ORFNames=BD7051;	
OC	uncultured organism.	
OX	NCBI_TaxID=155900;	
RN	[1]	
RP	NNUCLEOTIDE SEQUENCE.	
RK	PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;	
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,	
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,	
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.;	
RT	"Exploring nitrilase sequence space for enantioselective catalysis.";	
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).	
DR	EMBL; AY487559; AAC97506.1; -; Genomic DNA.	
DR	GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	
DR	GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . ; IEA.	
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.	
DR	InterPro; IPR000132; N/CN hydrtase.	
DR	InterPro; IPR003010; Ntlse/CNHdtase.	
DR	Pfam; PF00795; CN_hydrolase; 1.	
DR	PROSITE; PS50263; CN_HYDROLASE; 1.	
DR	PROSITE; PS00920; NITRIL_CHT_1; 1.	
DR	PROSITE; PS00921; NITRIL_CHT_2; 1.	
KW	Hydrolase.	
SQ	SEQUENCE 334 AA; 36445 MW; AOEB271DDFF3735E CRC64;	
	Query Match 60.9%; Score 1099.5; DB 2; Length 334;	
	Best Local Similarity 64.8%; Pred.No.5.8e-76;	
	Matches 210; Conservative 45; Mismatches 62; Indels 7; Gaps 3	
QY	8 YRGAAVQAAPVFLLDRTVEKAIGLIEQAQKVLRLLAAPETWIPGYEWIWLGA <sub>PAMGM</sub> 67 ::        :                     :                     :	
Db	8 FKAAVVQAAPFWLDLDRIVKTIALIEEAAGAKLIAFPETWIPGYWHIWGTPAWAI 67 :         :                     :                     :	
QY	68 R--FVQRYPENSLVRGSKWQALADARRHGMHVAVGYSERAGSLYMGCAIFGPDGLI 125                                 :                     :	
Db	68 SRGFVQRYFDNSLAYDSPAOIRIADAANKKITVLGLSEREGSLYSOMLIGPDET 127                                 :                     :	
QY	126 AARRKLKPThaERTVFGEGDSHLAVHDTA-IGRLGALCCWEHIQPLSKTYAMYAADQOVH 184         :                     :                     :         :	
Db	128 AKRRKLrPTHVERTVFGDGSHIAVHERADIGRLGALCWEHIQPLTKYAMAQNEDQVH 187         :                     :                     :         :	
QY	185 VASWPSSFSLYRGMAVALGPVENTAASQIYAveGGCYVLASCATVSPEMIKVLVDTPDKEM 244     :       :         :                     :     :       :	
Db	188 VAAMPFSFMEPFahalGWENTNAASKIYAVEGSCFVLGACAIVSQAMVDEMCDTEdkRa 247     :       :         :                     :     :       :	
QY	245 FLKAGGfAMIfgPDGRAlAEPLPETEEGLLVADIIdlGMtAlAKAAADPAghYSRpDVTR 304 :         :                     :                     :         :	





RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487546; AAR97493.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; NtIse/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
KM Hydrolase.  
SQ SEQUENCE 336 AA; 36221 MW; CC370C188F94CB72 CRC64;  
  
Query Match 59.2%; Score 1070; DB 2; Length 336;  
Best Local Similarity 60.2%; Pred. No. 1.1e-73;  
Matches 200; Conservative 53; Mismatches 75; Indels 4; Gaps 2;  
  
QY 1 MSEPMTKYRGAAVQAPVFLDLDRTEKAIIGLIEQAAKQDVRLIAFPETWIPGYFWIWL 60  
Db 3 ISHP--KFKAAVVOAGPAFLDLDDGVERAVSLIGQAAAEQAQLIAFPETWIPGYFWHTWL 60  
  
QY 61 GAPAWGMR--FVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSLYMGQAIF 118  
Db 61 GSPAWAMEKGFVQRYFDNALRHGSPQAEIRISGAABHKIMVSLGFAERDGGTLYIAQWLI 120  
  
QY 119 GPDGDLIAARRKLPKPTAERTVFEGEGDGHAVHDTAIGRLGALCCWEHIQPLSKYAMYA 178  
Db 121 GPDGQGISRRRLKPTVERTVFEGEGDSDLSVHDTALGRIGSLCCWEHIQPLSKYAMYA 180  
  
QY 179 ADEQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVD 238  
Db 181 QNEQIHIGAMPSPFSLYQPFANALSPENVIAVSRYAVEGQCFPLAPCATVSDAMIELTCD 240  
  
QY 239 TPDKEMFLKAGGFGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADPAGHYS 298  
Db 241 TPEKQGLIRAGGGGHAIFGPDGSLTPTVADTYEGLLYAEILDGVIStAKSAADPAGHYS 300  
  
QY 299 RPDVTRLILDRRPAQRVVTLDAFEPQNEDEK 330  
Db 301 RPDVTRLILNQTPSKRVQNMVLPLETYTEPEG 332  
  
RESULT 14  
Q6RWF7\_9ZZZZ PRELIMINARY; PRT; 335 AA.  
AC Q6RWF7;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORNames=BD5326;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487545; AAR97492.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.

DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; NtIse/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KM Hydrolase.  
SQ SEQUENCE 335 AA; 36902 MW; COABIDF5F963AC6A CRC64;  
  
Query Match 59.2%; Score 1068.5; DB 2; Length 335;  
Best Local Similarity 62.4%; Pred. No. 1.4e-73;  
Matches 209; Conservative 42; Mismatches 79; Indels 5; Gaps 2;  
  
QY 1 MSEPMTKYRGAAVQAPVFLDLDRTEKAIIGLIEQAAKQDVRLIAFPETWIPGYFWIWL 60  
Db 1 MSITHPKFKAAVVOAPVFLDLDDGSYKKAINLIDEAAAAGAKLIAFPETWIPGYFWIWL 60  
  
QY 61 GAPAW--GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSLYMGQAIF 118  
Db 61 GSPAWALGSGFVQRYFDNSLQYDSDPADRLREARRNSITVVLGHSERDGSGLYIAQWLI 120  
  
QY 119 GPDGDLIAARRKLPKPTAERTVFEGEGDGHAVHDTAIGRLGALCCWEHIQPLSKYAMYA 178  
Db 121 GPDGETIAQRRLKPTHTGERTVFEGEGDSDLVVHQTGLGRLGNWENILSLNKYVMYS 180  
  
QY 179 ADEQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVD 238  
Db 181 QHEQVHVASWPSFSSTYEPFAHALGYEVNNAISQVYAVEGGCYVLAPCSTISEEMIAELCD 240  
  
QY 239 TPDKEMFLKAGGFGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADPAGHYS 298  
Db 241 TPDKFELTHAGGGHAIIYGPDGRALCEKLPENQEGLLYAEILDGVIStAKSAMPVGHYS 300  
  
QY 299 RPDVTRLILDRRPAQRVVTLDAFEPQNEDEK 333  
Db 301 RPDVTVRLFNKTPAKRIEHFNL---FLDEQAGEEP 332  
  
RESULT 15  
Q5YUM5\_NOCFA PRELIMINARY; PRT; 336 AA.  
AC Q5YUM5;  
DT 25-OCT-2004 (TREMBlrel. 28, Created)  
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Putative nitrilase.  
GN OrderedlocusNames=nfa32690;  
OS Nocardia farcinica.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Nocardiaceae; Nocardia.  
OX NCBI\_TaxID=37329;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=IFM 10152;  
RX PubMed=15466710; DOI=10.1073/pnas.0406410101;  
RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,  
RA Shiba T., Hattori M.;  
RT "The complete genomic sequence of Nocardia farcinica IFM 10152.";   
Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).  
DR EMBL; AP006618; BAD58116.1; -; Genomic\_DNA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; NtIse/CNhydase.  
DR InterPro; IPR002332; PII\_GlnB\_UMP\_S.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN\_1.  
KM Complete proteome.  
SQ SEQUENCE 336 AA; 36091 MW; D24921911BD0C2BB CRC64;





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:05 ; Search time 135.766 Seconds  
(without alignments)  
1119.760 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806

Sequence: 1 MSEPMTKYRGAAYQAPVFL.....EDKGDAPALRVVAESAAMAAQ 346

Scoring table: BLOSUM62

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1806	100.0	346	4	AAE05488	Aae05488 Nitrilase
2	1806	100.0	346	7	ADC24117	Adc24117 Protein s
3	1806	100.0	346	8	ADH36218	Adh36218 Chemical
4	1806	100.0	346	8	ADG93917	Adg93917 Nitrilase
5	1806	100.0	346	8	ADI62516	Adi62516 Nitrilase
6	1806	100.0	346	8	ADI64637	Adi64637 Nitrilase
7	1464	81.1	341	7	ADC23767	Adc23767 Protein s
8	1464	81.1	341	8	ADH35868	Adh35868 Chemical
9	1464	81.1	341	8	ADG93569	Adg93569 Nitrilase
10	1464	81.1	341	8	ADI62166	Adi62166 Nitrilase
11	1464	81.1	341	8	ADI64287	Adi64287 Nitrilase
12	1209	66.9	348	7	ADC23749	Adc23749 Protein s
13	1209	66.9	348	8	ADH35850	Adh35850 Chemical
14	1209	66.9	348	8	ADG93551	Adg93551 Nitrilase
15	1209	66.9	348	8	ADI62148	Adi62148 Nitrilase
16	1209	66.9	348	8	ADI64269	Adi64269 Nitrilase
17	1165	64.5	353	7	ADC23945	Adc23945 Protein s
18	1165	64.5	353	8	ADH36046	Adh36046 Chemical
19	1165	64.5	353	8	ADG93747	Adg93747 Nitrilase
20	1165	64.5	353	8	ADI62344	Adi62344 Nitrilase
21	1165	64.5	353	8	ADI64465	Adi64465 Nitrilase
22	1155	64.0	353	7	ADC23965	Adc23965 Protein s
23	1155	64.0	353	8	ADH36066	Adh36066 Chemical
24	1155	64.0	353	8	ADG93767	Adg93767 Nitrilase

25	1155	64.0	353	8	ADI62364	Adi62364 Nitrilase
26	1155	64.0	353	8	ADI64485	Adi64485 Nitrilase
27	1148	63.6	333	7	ADC24083	Adc24083 Protein s
28	1148	63.6	333	8	ADH36184	Adh36184 Chemical
29	1148	63.6	333	8	ADG93883	Adg93883 Nitrilase
30	1148	63.6	333	8	ADI62482	Adi62482 Nitrilase
31	1148	63.6	333	8	ADI64603	Adi64603 Nitrilase
32	1132.5	62.7	358	7	ADC23903	Adc23903 Protein s
33	1132.5	62.7	358	8	ADH36004	Adh36004 Chemical
34	1132.5	62.7	358	8	ADG93705	Adg93705 Nitrilase
35	1132.5	62.7	358	8	ADI62302	Adi62302 Nitrilase
36	1132.5	62.7	358	8	ADI64423	Adi64423 Nitrilase
37	1130	62.6	354	7	ADC23835	Adc23835 Protein s
38	1130	62.6	354	8	ADH35936	Adh35936 Chemical
39	1130	62.6	354	8	ADG93637	Adg93637 Nitrilase
40	1130	62.6	354	8	ADI62234	Adi62234 Nitrilase
41	1130	62.6	354	8	ADI64355	Adi64355 Nitrilase
42	1125	62.3	345	7	ADC24067	Adc24067 Protein s
43	1125	62.3	345	8	ADH36168	Adh36168 Chemical
44	1125	62.3	345	8	ADG93868	Adg93868 Nitrilase
45	1125	62.3	345	8	ADI62466	Adi62466 Nitrilase

ALIGNMENTS

RESULT 1	
AAE05488	AAE05488 standard; protein; 346 AA.
XX	
AC	AAE05488;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Nitrilase protein #1.
XX	
KW	Nitrilase; enantiomer; alpha-substituted carboxylic acid;
KW	stereoselective; hydrolysis; amino nitrile; cyanohydrin.
XX	
OS	Unidentified.
XX	
PN	WO200148175-A2.
XX	
PD	05-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US035555.
XX	
PR	29-DEC-1999; 99US-0173609P.
PR	07-DEC-2000; 2000US-0254414P.
XX	
PA	(DIVE-) DIVERSA CORP.
PA	(MADD/) MADDEN D.
XX	
PI	Madden M, Weiner DP, Chaplin JA;
XX	
DR	WPI; 2001-465211/50.
DR	N-PSDB; AAD11166.
XX	
PT	Producing alpha-substituted carboxylic acid enantiomers by contacting
PT	aldehyde or ketone with cyanide, ammonia compound or its salt or amine
PT	and hydrolyzing the resulting amino nitrile or cyanohydrin with
PT	Nitrilase.
XX	
PS	Claim 29; Page 84-85; 87pp; English.
XX	
CC	The present invention relates to methods for producing an
CC	enantiomerically pure alpha-substituted carboxylic acid. The method
CC	involves contacting an aldehyde or ketone with a cyanide containing
CC	compound and an ammonia-containing compound or an ammonium salt or amine,
CC	and stereoselectively hydrolysing the resulting amino nitrile or
CC	cyanohydrin intermediate with a nitrilase or a polypeptide having
CC	nitrilase activity. The present sequence is a nitrilase protein #1
XX	

SQ Sequence 346 AA;
Query Match 100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 5.7e-172;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIML 60
DB 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIML 60
QY 61 GAPAWGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
DB 61 GAPAWGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
DB 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
DB 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300
DB 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300
QY 301 DVTIRLLDRRPAQRVVTLDAAFEPOKEDKGDAPALRVVAESAAAAQ 346
DB 301 DVTIRLLDRRPAQRVVTLDAAFEPOKEDKGDAPALRVVAESAAAAQ 346

RESULT 2
ADC24117
ID ADC24117 standard; protein; 346 AA.
XX
AC ADC24117;
XX
DT 18-DEC-2003 (first entry)
XX
DE Protein sequence (SeqID 384) exhibiting nitrilase activity.
XX
KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;
KW enantiomer; chiral medicine.
XX
OS Unidentified.
XX
PN WO2003000840-A2.
XX
PD 03-JAN-2003.
XX
PF 15-MAY-2002; 2002WO-US015983.
XX
PR 21-JUN-2001; 2001US-0300189P.
PR 30-JUL-2001; 2001US-0309006P.
PR 22-JAN-2002; 2002US-0351336P.
XX
PA (DIVE-) DIVERSA CORP.
PA (MADD/) MADDEN D.
XX
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;
PI Short JM, Burk M;
XX
DR WPI; 2003-201417/19.
DR N-PSDB; ADC24116.
XX
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl
PT lactic acid derivative and for producing pharmaceutical composition, and
PT food additive.
XX
PS Claim 40; SEQ ID NO 384; 560bp; English.
XX
SC This invention relates to nitrilases and the nucleic acids that encode

CC these enzymes thereof. Specifically, it refers to polypeptides that
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse
CC nitriles or cyanohydrins into their corresponding carboxylic acids and
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in
CC the synthesis of enantiomerically pure aromatic and aliphatic amino
CC acids, as well as hydroxy acids, which are important for the development
CC of chiral medicines. Furthermore, the present invention describes
CC nitrilases, isolated from mesophilic microorganisms, that have improved
CC activity and stability at increased pH and temperature. They are also
CC inexpensive, efficient catalysts, have broad substrate specificity and
CC are capable of chiral differentiation. This polypeptide is a protein
CC sequence that exhibits nitrilase activity of the invention.
XX
SQ Sequence 346 AA;
Query Match 100.0%; Score 1806; DB 7; Length 346;
Best Local Similarity 100.0%; Pred. No. 5.7e-172;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIML 60
DB 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIML 60
QY 61 GAPAWGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
DB 61 GAPAWGMRVQRYFENSLVRGSKQWQALADAAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
DB 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
DB 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300
DB 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRP 300
QY 301 DVTIRLLDRRPAQRVVTLDAAFEPOKEDKGDAPALRVVAESAAAAQ 346
DB 301 DVTIRLLDRRPAQRVVTLDAAFEPOKEDKGDAPALRVVAESAAAAQ 346

RESULT 3
ADH36218
ID ADH36218 standard; protein; 346 AA.
XX
AC ADH36218;
XX
DT 11-MAR-2004 (first entry)
XX
DE Chemical process monitoring-related nitrilase protein sequence SeqID384.
XX
KW chemical process monitoring; biochemical process monitoring; cyanide;
KW high throughput system; enzyme.
XX
OS Unidentified.
XX
PN WO2003098187-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003WO-US015639.
XX
PR 15-MAY-2002; 2002US-0380737P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI Mcquaid J, Stege J;
XX
DR WPI; 2004-142708/14.

DR N-PSDB; ADH36217.

XX Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.

XX PS Claim 73; SEQ ID NO 384; 277pp; English.

XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.

XX SQ Sequence 346 AA;

Query Match 100.0%; Score 1806; DB 8; Length 346;  
Best Local Similarity 100.0%; Pred. No. 5.7e-172;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60  
Db 1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60

OY 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGP 120  
Db 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGP 120

OY 121 DGDLLIARRRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

OY 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240  
Db 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

OY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHSRP 300

OY 301 DVTTRLDDRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346  
Db 301 DVTTRLDDRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346

RESULT 4  
ADG93917  
ID ADG93917 standard; protein; 346 AA.

XX AC ADG93917;

XX DT 11-MAR-2004 (first entry)

DE Nitrilase enzyme amino acid sequence SegID384.

KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; enzyme.

XX OS Unidentified.

XX PN WO2003097810-A2.

XX PD 27-NOV-2003.

XX PF 15-MAY-2003; 2003WO-US015712.

XX PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

XX PA (DIVE-) DIVERSA CORP.

XX PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

XX DR WPI; 2004-090637/09.

XX DR N-PSDB; ADG93916.

XX PT New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.

XX PS Claim 44; SEQ ID NO 384; 295pp; English.

XX This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.

XX SQ Sequence 346 AA;

Query Match 100.0%; Score 1806; DB 8; Length 346;  
Best Local Similarity 100.0%; Pred. No. 5.7e-172;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60  
Db 1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60

OY 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGP 120  
Db 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAIFGP 120

OY 121 DGDLLIARRRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

OY 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240  
Db 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

OY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHSRP 300

OY 301 DVTTRLDDRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346  
Db 301 DVTTRLDDRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346

RESULT 5  
ADI62516  
ID ADI62516 standard; protein; 346 AA.

XX AC ADI62516;

XX DT 22-APR-2004 (first entry)

DE Nitrilase polypeptide #192.

KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;

KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;

KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;

KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipemic; enzyme.

XX OS Unidentified.



```
XX PN WO2003106415-A2.
XX PD 24-DEC-2003.
XX PF 13-JUN-2003; 2003WO-US018840.
XX PR 13-JUN-2002; 2002US-0389317P.
XX PR 28-JUN-2002; 2002US-0392944P.
XX PA (DIVE-) DIVERSA CORP.
XX PI Burk M, Desantis G, Morgan B, Zhu Z;
XX DR WPI; 2004-090821/09.
XX DR N-PSDB; ADI62515.
XX PT Preparation of atorvastatin comprises catalytic conversion of 3-
XX PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
XX PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX PT acid and forming atorvastatin.
XX PS Claim 41; SEQ ID NO 384; 253pp; English.
XX CC The present invention relates to a method for preparing an atorvastatin
XX CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
XX CC cyano-3-hydroxybutyric acid). The method comprises optionally converting
XX CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
XX CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
XX CC hydroxybutyric acid with a polypeptide having nitrilase activity,
XX CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
XX CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
XX CC method involves whole cell processes, cell lysate process, "one pot"
XX CC processes, and "multi-pot" processes using a variety of parameters.
XX CC Atorvastatin is used, in conjunction with dietary restriction, in the
XX CC management of hyperlipidaemia, including hypercholesterolaemia, mixed
XX CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present
XX CC sequence represents a nitrilase polypeptide obtained from an
XX CC environmental sample.
SQ Sequence 346 AA;
Query Match 100.0%; Score 1806; DB 8; Length 346;
Best Local Similarity 100.0%; Pred. No. 5.7e-172;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60
DB 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60
QY 61 GAPAWGMRVFQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQAIIFGP 120
DB 61 GAPAWGMRVFQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQAIIFGP 120
QY 121 DGDLLAARRKLKPTHAERTVFGEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
DB 121 DGDLLAARRKLKPTHAERTVFGEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
DB 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMITALAKAAADPAGHYSRP 300
DB 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMITALAKAAADPAGHYSRP 300
QY 301 DVTRLLLDRRPAQRVVTLDAAFEPPQNEKGDAPALRVVAESAAAAQ 346
DB 301 DVTRLLLDRRPAQRVVTLDAAFEPPQNEKGDAPALRVVAESAAAAQ 346
RESULT 6
, ADI64637
```

```
ID ADI64637 standard; protein; 346 AA.
XX AC ADI64637;
XX DT 22-APR-2004 (first entry)
XX DE Nitrilase seq id 193.
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
KW carboxylic acid; cyanohydrin moiety hydrolysis;
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
KW (R)-phenyl lactic acid derivative; & enantiomeric excess;
KW & diastomeric excess; food additive; drug intermediate; nitrilase.
XX Unidentified.
XX OS US2004014195-A1.
XX PN 22-JAN-2004.
XX PD 15-MAY-2003; 2003US-00440523.
XX PF 29-DEC-1999; 99US-0173609P.
XX PR 07-DEC-2000; 2000US-0254414P.
XX PR 28-DEC-2000; 2000US-00751299.
XX PR 31-JUN-2001; 2001US-0300189P.
XX PR 30-JUL-2001; 2001US-0309006P.
XX PR 22-JAN-2002; 2002US-0351336P.
XX PR 15-MAY-2002; 2002US-00146772.
XX PR 09-SEP-2002; 2002US-00241742.
XX PA (DIVE-) DIVERSA CORP.
XX PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX DR WPI; 2004-121569/12.
XX DR N-PSDB; ADI64636.
XX PT Novel isolated or recombinant polypeptide having nitrilase activity,
XX PT useful in production of food additives.
XX PS Claim 46; SEQ ID NO 384; 105pp; English.
XX CC The invention describes an isolated or recombinant polypeptide (I)
XX CC comprising amino acids having a sequence at least 50 % identical to a
XX CC sequence (SI) available in electronic form (EC) from the following web
XX CC site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its
XX CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
XX CC residue 60 glutamic acid, at residue 111 Ser, their combinations or
XX CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
XX CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
XX CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
XX CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
XX CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
XX CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
XX CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
XX CC and for identifying a modified compound. The inventive method is useful
XX CC for monitoring or determining & enantiomeric excess or & diastomeric
XX CC excess. (I) is useful in the production of food additives and drug
XX CC intermediates. This is the amino acid sequence of a nitrilase of the
XX CC invention.
SQ Sequence 346 AA;
Query Match 100.0%; Score 1806; DB 8; Length 346;
Best Local Similarity 100.0%; Pred. No. 5.7e-172;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60
DB 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIWL 60
```



CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.

Query Match	81.1%;	Score 1464;	DB 8;	Length 341;
Best Local Similarity	81.7%;	Pred. No. 1.1e-137;		
Matches 272;	Conservative	28;	Mismatches 33;	Indels 0;
				Gaps 0;

QY	1	MSEPMTKYRGAAVQAA	PVFLDLDRTEVKAIGLIEQAAKODVRLIAFPETWIPGYPFWTWL	60
		:::	:::	:::
Db	1	MLSPVTQYRAAAVQAA	PSFLDLDRTEVKTIAIEQAAEQDVRLLIAFPETWIPGYPLWTWL	60
QY	61	GAPAWGMR	FVQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAGGSLYMGAIFGP	120
		:::	:::	:::
Db	61	GSPAWGMR	FVQRYFENSLVRGSKQWNAIADARRHRMTVVVGFSEERAGGSLYMGAIFGP	120
QY	121	DGDLLIARRKLKPT	HAERTVFGEGDGHVAHDTAIGRLGALCCWEHIQPLSKYAMYAAD	180
		:::	:::	:::
Db	121	EGELIARRKLKPT	HAERTVFGEGDGHVAHYETGVGRIGALCCWEHIQPLSKYAMYAAN	180
QY	181	EQVHVASWPSFSLY	RGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP	240
		:::	:::	:::
Db	181	EQVHVASWPCFSLY	RGMAYALGPEVNTAASQYVAVEGGCYVLASCLVVTPEILKVLIDTP	240
QY	241	DKEMFLKAGGGFAMI	FGPDGRALAEPLPETHEGLLVADIDLGMIALAKAAADPAGHYSRP	300
		:::	:::	:::
Db	241	DKEPULLLAGGGF	SMIFGPDGRALAQPLPETHEGLVTAETIDLGAIALAKAAADPAGHYARP	300
QY	301	DVTRLLLLDRRPAQ	RVTTLDAFEPQNEDEKGDAP	333
		:::	:::	:::
Db	301	DVTRLLLNPRPAAR	VEALGPRFEVVQSEQAEP	333

RESULT 9  
ADG93569  
ID ADG93569 standard; protein; 341 AA.

DT 11-MAR-2004 (first entry)

DE Nitrilase enzyme amino acid sequence SegID34.

KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; enzyme.

PN WO2003097810-A2.

PD 27-NOV-2003.

PF 15-MAY-2003; 2003WO-US015712.

PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.

PA (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

DR WPI; 2004-090637/09.  
DR N-PSDB; ADG93568.

PT	New isolated or recombinant nucleic acid encoding a polypeptide having
PT	nitrilase activity, useful for screening enantioselective transformation
XX	
PS	Claim 44; SEQ ID NO 34; 295pp; English.

CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.

Query Match	81.1%;	Score 1464;	DB 8;	Length 341;
Best Local Similarity	81.7%;	Pred. No. 1.1e-137;		
Matches 272;	Conservative	28;	Mismatches 33;	Indels 0;
			Gaps	0;

QY	I MSEPMTKYRGAAVQAAPFVLDDRTVEKAIGLIEQAOKODVRLLIAFPETWIPGYPFMTWL	60
Db	1 MISPVTQYRAAAVQAAPSFLDLDRTVEKTAIIIEQAEEQDVRLIAFPETWIPGYPLTWL	60
QY	61 GAPAWGMRVQRYFFENSLVRGSKOMALADARRHGHMVAGYSERAGGSLYMGQAIFGP	120C
Db	61 GSPAWGMRFVQRYFFENSLVRGSKOMNAIADARRHRMTVVVGFSERAGGSLYMGQAIFGP	120C
QY	121 DGDLLIARRKLKPHTHAERTVFEGEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD	180C
Db	121 EGELIARRKLKPHTHAERTVFEGEDGSHLAVYETGVGRIGALCCWEHIQPLSKYAMYAAN	180C
QY	181 EQVHVASWPSSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP	240C
Db	181 EQVHVASWPCFSFLYRGMAYALGPEVNTAASQVYAVEGGCYVLASCLVWTPEILKVLIDTP	240C
QY	241 DKEMFLKAGGGFAMIFGPDGRALAEPUPETEGLLVADIDLGMTALAKAAAADPAGHYSRP	300C
Db	241 DKEPLLLAGGGFSMIFGPDGRALAOPLPETEGLVTAETIDLGAIALAKAAAADPAGHYARP	300C
QY	301 DVTRLILLDRRPAQRVVTLDAAFEPONEDEKGDAP	333
Db	301 DVTRLILNPRPARVEALGPRFEVVOSEQAEPP	333

RESULT 10	
ADI62166	
ID	ADI62166 standard; protein; 341 AA

DT 22-APR-2004 (first entry)

DE Nitrilase polypeptide #17.

KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipaeic; enzyme.

PN WO2003106415-A2.

PD 24-DEC-2003

PF 13-JUN-2003; 2003WO-US018840.

PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.

XX (DIVE-) DIVERSA CORP.  
PA  
XX Burk M, Desantis G, Morgan B, Zhu Z;  
PI  
XX WPI; 2004-090821/09.  
DR  
DR N-PSDB; ADI62165.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 41; SEQ ID NO 34; 253pp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate processes, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrilase polypeptide obtained from an  
CC environmental sample.  
XX  
SQ Sequence 341 AA;  
  
Query Match 81.1%; Score 1464; DB 8; Length 341;  
Best Local Similarity 81.7%; Pred. No. 1.1e-137;  
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;  
  
QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVKAIGLIEQAAKQDVRLIAPETWIPGYFWIWL 60  
DB 1 MLSPTQYRAAAVQAAPSFLLDLRTVEKTIATIEQAAEQDVRLIAPETWIPGYPLMIWL 60  
  
QY 61 GAPAWGMRFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSLYMGQAIFGP 120  
DB 61 GSPAWGMRFVQRYFENSLVRGSKQWALADARRHMTVVVGFSEKAGSLYMGQAIFGP 120  
  
QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
DB 121 EGELIAARRKLPKPTHAERTVFGEGDGSHLAVYETGVGRIGALCWEHIQPLSKYAMYAAN 180  
  
QY 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240  
DB 181 EQVHVASWPCFSLYRGMAAYALGPEVNTAASQVYAVEGGCYVLASCLVTPPEIKVLIDTP 240  
  
QY 241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRP 300  
DB 241 DKEPLLLAGGGSFMIFGPDGRALAOPLPETEEGLVTAEIDLGAIALAKAAADPAGHYARP 300  
  
QY 301 DVTIRLLDRRPAQRVVTLDAAFEQNEDEKGDAP 333  
DB 301 DVTIRLLNRPAPARVEALGPRFEVVGSEQAEP 333  
  
RESULT 11  
ADI64287  
ID ADI64287 standard; protein; 341 AA.  
XX  
AC ADI64287;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase seq id 18.  
XX  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;

KW carboxylic acid; cyanohydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KW (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KW & diastomeric excess; food additive; drug intermediate; nitrilase.  
XX  
OS Unidentified.  
XX  
PN US2004014195-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 15-MAY-2003; 2003US-00440523.  
XX  
PR 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-121569/12.  
DR N-PSDB; ADI64286.  
XX  
PT Novel isolated or recombinant polypeptide having nitrilase activity,  
PT useful in production of food additives.  
XX  
PS Claim 46; SEQ ID NO 34; 105pp; English.  
XX  
CC The invention describes an isolated or recombinant polypeptide (I)  
CC comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspto.gov/sequence.html?docid=2004014195, or its  
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining & enantiomeric excess or & diastomeric  
CC excess. (I) is useful in the production of food additives and drug  
CC intermediates. This is the amino acid sequence of a nitrilase of the  
CC invention.  
XX  
SQ Sequence 341 AA;  
  
Query Match 81.1%; Score 1464; DB 8; Length 341;  
Best Local Similarity 81.7%; Pred. No. 1.1e-137;  
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;  
  
QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEVKAIGLIEQAAKQDVRLIAPETWIPGYFWIWL 60  
DB 1 MLSPTQYRAAAVQAAPSFLLDLRTVEKTIATIEQAAEQDVRLIAPETWIPGYPLMIWL 60  
  
QY 61 GAPAWGMRFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSLYMGQAIFGP 120  
DB 61 GSPAWGMRFVQRYFENSLVRGSKQWALADARRHMTVVVGFSEKAGSLYMGQAIFGP 120  
  
QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
DB 121 EGELIAARRKLPKPTHAERTVFGEGDGSHLAVYETGVGRIGALCWEHIQPLSKYAMYAAN 180





CC invention.  
XX  
SQ Sequence 348 AA;

Query Match 66.9%; Score 1209; DB 8; Length 348;  
Best Local Similarity 67.1%; Pred. No. 3.9e-112;  
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

QY 4 PMTK-YRGAAVQAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAPPETWIPGYPFWIMGA 62  
| : : | | | | | | | | | | : : | | | | | : | | | | | | | | | : | | | | | :  
Db 2 PTKQFRVAAVQAPVFLDLLEGALISKGISLIEBAASNGAKLIAFPETWIPGYPWMWILDS 61  
  
QY 63 PAMGMRPVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAI FGPDG 122  
| | | | | | | | | | :  
Db 62 PAMGMRPVQRYFDNSLMLGSEQAKRMNQAAANNKIYVMGYSERGSGSLYMGQSIINDKG 121  
  
QY 123 DLIAARRKLKPTHAERTVFEGEGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQ 182  
: | : | | | | | | | | | | : : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 122 ETIFTRRLKPTHTVERTVFEGEGSHLCVMDTEIGRVGAMCCWEHLQPLSKYAMYSQDEQ 181  
  
QY 183 VHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTDPK 242  
: | : | | | | | | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 182 IHIASWPSFSLYRGAAAYALGPFLNNNAASQMYAAEGQCFVLAPCATVSKEMIEMLIDDPK 241  
  
QY 243 EMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRPDV 302  
| | | | | | | | | | :  
Db 242 EPLLEGGGFTWYGPDPGRPLAKPLPENEGLLYADIDLGMISMAKAAADPAGHYARPDV 301  
  
QY 303 TRLLDRRPAQRVVTLDAAFEPPQNEKGDAPALRVVAESAAAA 345  
| | | | : | | | : : | | | : : : | |  
Db 302 TRLFNSAPANRVEYINPASGP-TESLKMGMQMEAEQOKAA 343

RESULT 14  
ADG93551  
ID ADG93551 standard; protein; 348 AA.  
XX  
AC ADG93551;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitrilase enzyme amino acid sequence SeqID16.  
XX  
KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003097810-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-090637/09.  
DR N-PSDB; ADG93550.  
XX

PT New isolated or recombinant nucleic acid encoding a polypeptide having  
nitrilase activity, useful for screening enantioselective transformation.

PS Claim 44; SEQ ID NO 16; 295pp; English.

XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrlase activity. Nitrlase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated

CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrlase enzyme of the invention.

XX  
SQ Sequence 348 AA;

Query Match 66.9%; Score 1209; DB 8; Length 348;  
Best Local Similarity 67.1%; Pred. No. 3.9e-112;  
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

QY 4 PMTK-YRGAAVQAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAPPETWIPGYPFWIMGA 62  
| : : | | | | | | | | | | : : | | | | | : | | | | | | | | | : | | | | | :  
Db 2 PTKQFRVAAVQAPVFLDLLEGALISKGISLIEBAASNGAKLIAFPETWIPGYPWMWILDS 61  
  
QY 63 PAMGMRPVQRYFENSLVRGSKQWQALADAAARRHGMHVAVGYSERAGGSLYMGQAI FGPDG 122  
| | | | | | | | | | :  
Db 62 PAMGMRPVQRYFDNSLMLGSEQAKRMNQAAANNKIYVMGYSERGSGSLYMGQSIINDKG 121  
  
QY 123 DLIAARRKLKPTHAERTVFEGEGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQ 182  
: | : | | | | | | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 122 ETIFTRRLKPTHTVERTVFEGEGSHLCVMDTEIGRVGAMCCWEHLQPLSKYAMYSQDEQ 181  
  
QY 183 VHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTDPK 242  
: | : | | | | | | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :  
Db 182 IHIASWPSFSLYRGAAAYALGPFLNNNAASQMYAAEGQCFVLAPCATVSKEMIEMLIDDPK 241  
  
QY 243 EMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRPDV 302  
| | | | : | | | : : | | | : : : | |  
Db 242 EPLLEGGGFTWYGPDPGRPLAKPLPENEGLLYADIDLGMISMAKAAADPAGHYARPDV 301  
  
QY 303 TRLLDRRPAQRVVTLDAAFEPPQNEKGDAPALRVVAESAAAA 345  
| | | | : | | | : : | | | : : : | |  
Db 302 TRLFNSAPANRVEYINPASGP-TESLKMGMQMEAEQOKAA 343

RESULT 15  
ADI62148  
ID ADI62148 standard; protein; 348 AA.  
XX  
AC ADI62148;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrlase polypeptide #8.  
XX  
KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipaeamic; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.

PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
DR WPI; 2004-090821/09.  
DR N-PSDB; ADI62147.  
XX

PT Preparation of atorvastatin comprises catalytic conversion of 3-



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:00:30 ; Search time 34.448 Seconds  
(without alignments)  
830.404 Million cell updates/sec

Title: US-09-751-299-2  
Perfect score: 1806  
Sequence: 1 MSEPMTKYRGAAVQAAPVFL.....EDKGDAPALRVVAESAAAAQ 346

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	933	51.7	344	1	US-08-690-493-1 Sequence 1, Appli
2	869	48.1	356	2	US-09-806-876A-2 Sequence 2, Appli
3	807	44.7	369	2	US-09-823-373-5 Sequence 5, Appli
4	807	44.7	369	2	US-09-823-373-14 Sequence 14, Appli
5	592.5	32.8	354	1	US-08-447-702-5 Sequence 5, Appli
6	592.5	32.8	354	1	US-08-465-615-5 Sequence 5, Appli
7	176.5	9.8	325	2	US-09-543-681A-7440 Sequence 2, Appli
8	150	8.3	315	2	US-10-105-294B-2 Sequence 5160, Ap
9	145.5	8.1	175	2	US-09-621-976-5160 Sequence 10443, A
10	144.5	8.0	337	2	US-09-489-039A-10443 Sequence 21992, A
11	135.5	7.5	302	2	US-09-252-991A-21992 Sequence 7670, Ap
12	131.5	7.3	152	2	US-09-621-976-7670 Sequence 14601, A
13	126	7.0	267	2	US-09-902-540-14601 Sequence 1, Appli
14	124.5	6.9	501	2	US-09-863-339A-1 Sequence 24644, A
15	117.5	6.5	349	2	US-09-252-991A-24644 Sequence 2, Appli
16	117	6.5	370	2	US-09-863-339A-2 Sequence 12181, A
17	116	6.4	664	2	US-09-902-540-12181 Sequence 28974, A
18	114.5	6.3	399	2	US-09-252-991A-28974 Sequence 4728, Ap
19	109	6.0	541	2	US-09-328-352-4728 Sequence 6533, Ap
20	108	6.0	513	2	US-09-949-016-6533 Sequence 7160, Ap
21	108	6.0	517	2	US-09-949-016-7160 Sequence 5, Appli
22	104.5	5.8	409	1	US-08-809-740A-5 Sequence 5130, Ap
23	101.5	5.6	291	2	US-09-583-110-5130 Sequence 4874, Ap
24	101.5	5.6	294	2	US-09-107-433-4874 Sequence 22107, A
25	101.5	5.6	360	2	US-09-252-991A-22107 Sequence 16772, A
26	101	5.6	339	2	US-09-902-540-16772 Sequence 26229, A
27	101	5.6	633	2	US-09-252-991A-26229 Sequence 26, Appli

28	100.5	5.6	540	2	US-09-252-991A-23300 Sequence 23300, A
29	100	5.5	648	2	US-09-902-540-14001 Sequence 14001, A
30	100	5.5	759	2	US-09-252-991A-33071 Sequence 33071, A
31	98	5.4	311	1	US-07-917-111-5 Sequence 5, Appli
32	98	5.4	311	1	US-07-917-111-6 Sequence 6, Appli
33	98	5.4	311	1	US-08-479-638-5 Sequence 5, Appli
34	98	5.4	311	1	US-08-479-638-6 Sequence 6, Appli
35	98	5.4	311	1	US-08-294-871A-70 Sequence 70, Appli
36	98	5.4	311	2	US-08-876-398A-70 Sequence 27055, A
37	97.5	5.4	346	2	US-09-252-991A-27055 Sequence 11251, A
38	96	5.3	1990	2	US-09-902-540-11251 Sequence 23838, A
39	95	5.3	514	2	US-09-252-991A-22124 Sequence 1, Appli
40	92.5	5.1	405	2	US-09-252-991A-23838 Sequence 11585, A
41	92.5	5.1	4563	2	US-09-108-006C-1 Sequence 32525, A
42	92	5.1	630	2	US-09-902-540-11585 Sequence 27198, A
43	92	5.1	696	2	US-09-252-991A-32525 Sequence 26, Appli
44	91.5	5.1	401	2	US-09-252-991A-27198 Sequence 26, Appli
45	91.5	5.1	415	2	US-08-861-774E-26 Sequence 26, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-1  
Sequence 1, Application US/08690493  
Patent No. 5872000  
GENERAL INFORMATION:  
APPLICANT: Yu, Fujio  
TITLE OF INVENTION: No. 5872000e1 Nitrilase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steinberg, Raikin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,493  
FILING DATE: 31 JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 213061/1995  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 3821005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 768-3800  
TELEFAX: (212) 382-2124  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acid residues  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gordona terrae  
STRAIN: MA-1  
CELL TYPE: unicellular organism  
FEATURE:  
OTHER INFORMATION: Xaa is Met or a deletion  
US-08-690-493-1  
Query Match 51.7%; Score 933; DB 1; Length 344;  
Best Local Similarity 55.2%; Pred. No. 7.5e-98;  
Matches 170; Conservative 52; Mismatches 86; Indels 0; Gaps 0;



QY 8 YRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGAPAWGM 67  
Db 9 FKAAVTQAEPVWFDLSATVDKTIATLVEEASRAGADLIAFPETWIPGYPFWIMLDSVAMOS 68  
QY 68 RFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLYMGQAIFGPDGLIAA 127  
Db 69 QYFIRYPQNSLDLDGSEFAIRREARKNDIAITMGFSEKRGHSLYMGQAVIERDGVVVRT 128  
QY 128 RRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQVHVAS 187  
Db 129 RRKLKPTHVERTLFEGDGSDLVVDQTSLGRVGSLLCCWEHIQPLTKYAMYSQHEQIHIAA 188  
QY 188 WPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPDKEMFLK 247  
Db 189 WPSFSIFPGAVYALGPEVNTAASQOYAVEGQTYVLAPCAVIGDAGWEAFADTEEKROLIH 248  
QY 248 AGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDVTRLIL 307  
Db 249 KGGGYARITYGPDGRSLAEPLADNDEGILYADIDLSAIILAKNPADPVGHYSRPDVRLIGF 308  
QY 308 DRPAPQRY 315  
Db 309 NKAPOPKV 316

RESULT 2  
US-09-806-876A-2  
; Sequence 2, Application US/09806876A  
; Patent No. 6869783  
; GENERAL INFORMATION:  
; APPLICANT: Ress-Loeschke, Marion  
; APPLICANT: Friedrich, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: A process for preparing chiral carboxylic acids  
; TITLE OF INVENTION: from nitriles using a nitrilase or microorganisms  
; TITLE OF INVENTION: which comprise a gene for the nitrilase  
; FILE REFERENCE: 49462  
; CURRENT APPLICATION NUMBER: US/09/806, 876A  
; CURRENT FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: Germany/19848129.2  
; PRIOR FILING DATE: 1998-10-19  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: WordPerfect version 6.1  
; SEQ ID NO 2  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-09-806-876A-2

Query Match 48.1%; Score 869; DB 2; Length 356;  
Best Local Similarity 50.3%; Pred. No. 1.7e-90;  
Matches 171; Conservative 54; Mismatches 105; Indels 10; Gaps 2;  
QY 9 RGAAVQAAPVFLDLDRIVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGAPAWGM 68  
Db 8 RAAAVQAASPNYDLATGVDTITELARQARDEGCDLIYGETWLPGYPFHWLGGAPAWSLK 67  
QY 69 FVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLYMGQAIFGPDGLIAA 128  
Db 68 YSARYYANSLDLSAEFORIAQARTLGIFALGYSEKSGSLYLQCLIDDKGEMWMSR 127  
QY 129 RKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQVHVAS 188  
Db 128 RKLKPTHVERTLFEGYARDLIVSDTELGRVGCWEHLSPLSKYALYSQHEAIIHIAW 187  
QY 189 PSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPDKEMFLK 248  
Db 188 PSFSLYSEQAHALSAKVMAASQIYSVEGQCFITIASSVVTQETLDMLEVEGHNAPLLKV 247  
QY 249 GGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDVTRLIL 308  
Db 248 GGGSSMIFAPDGRTLAPYLPDAEGLIADLNMEETAFAKAINDPVGHYSKPEATRLVLD 307

QY 309 ---RRPAQRYVTILDAAFEPOEDKGDAPALRVVAESAAA 345  
Db 308 LGHRDPMTRV-----HSKSVTREAPEQGVQSKIASVA 340

RESULT 3  
US-09-823-373-5  
; Sequence 5, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; TITLE OF INVENTION: Acidovorax Facilis 72w  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; PRIOR FILING DATE: 2000-03-31  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Acidovorax facilis  
US-09-823-373-5

Query Match 44.7%; Score 807; DB 2; Length 369;  
Best Local Similarity 46.6%; Pred. No. 2.2e-83;  
Matches 160; Conservative 52; Mismatches 127; Indels 4; Gaps 2;  
QY 6 TKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGAPAW 65  
Db 6 SKFLATVQAEPWLDADATIDKISIGIEBAQKASLIAFPVEFIPGYPYMAWLDVKY 65  
QY 66 GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLYMGQAIFGPDGLI 125  
Db 66 SLFTSRHYENSLLECGDDRMRLQLAARRNKIALVMGYSEREAGSRYLQVFIERGEIV 125  
QY 126 AARRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADEQVHV 185  
Db 126 ANRRKLKPTHVERTIYGEENGTDPLTHDFAFGVGLNCWEHFQPLSKFMVYSLGEQVHV 185  
QY 186 ASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPDKEMF 245  
Db 186 ASWPAMSPLQPDVFOQLSEANATVTRSYAIEGQTFVLCSTQVIGPSALETFCLNDEQRAL 245  
QY 246 LKAGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDVTRL 305  
Db 246 LPQCGGARITYGPDGSELAKPLAEDAEGILYAEIDLEQILLAKAGADPVGHYSRPDVLSV 305  
QY 306 LIDRR---PAQRYVTILDAAFEPOEDKGDAPALRVVAESAAA 345  
Db 306 QFDPNRNHTPVHR-IGIDGRILDVNTRSRVENFRRLQAAEOERQA 347

RESULT 4  
US-09-823-373-14  
; Sequence 14, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; FILE REFERENCE: BC-1032 US NA



```
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,615
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,588
; FILING DATE: 10-FEB-1994
; APPLICATION NUMBER: FR 9209882
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 003025-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-615-5

Query Match      32.8%; Score 592.5; DB 1; Length 354;
Best Local Similarity 39.4%; Pred. No. 7.3e-59;
Matches 137; Conservative 51; Mismatches 133; Indels 27; Gaps 7;

QY 11 AAVOAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLG----APAW 65
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 10 AAVOAAPVFMNLEATVDKTKLIAEASMGAKVIGFPEAFIPGYPYWIMTSMNDFGMMW 69
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 66 GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSERAGGSLYMGQAIFGPDGLI 125
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 70 AV-----LFKNAIEIPSKVEVQISDAAKNGVYCVSVSEKDNASLYLTQLWFDPNGLI 124
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 126 AARRKIKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQYHV 185
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 125 GHRKRFKPTSSERAVWGDGSMAPVFEKTEYGNLGLQCWEHALPLNIAAMGSLNEQYHV 184
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 186 ASWPSF-----SLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLYD 238
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 185 ASWPAFVPKGAIVSSRVSSVCASTNAMHQIISQFYAISNQYVIMSTNLVGQDMIDMGK 244
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 239 TPDKEMFLKAGGFGAMIFGPDGRALAEPLPETEGLVADIDLGMIALAKAADPAQHYH 298
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 245 DEFSKNFLPLGSGNTAISNTGEILAS-IPQDAEGIAVAEIDLNQIITYGKWLDPAGHYS 303
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 299 RPDVTRLILDRR---PAQRV---VTLDAAFEPOQNEKGD---APALRV 337
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 304 TPGFLSLTFDQSEHVPVKKIGQTNHFISYEDLHEDKMDMLTIPPRV 351

RESULT 7
US-09-543-681A-7440
; Sequence 7440, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7440
; LENGTH: 325
```

```
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-7440

Query Match      9.8%; Score 176.5; DB 2; Length 325;
Best Local Similarity 24.7%; Pred. No. 2.6e-11;
Matches 85; Conservative 50; Mismatches 134; Indels 75; Gaps 15;

QY 4 PMTKYGAVOAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGAP 63
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 5 PLTLRVASVQLQHKANDKDYNLAKIHQFIEMAASEKNLLVFPEMCITGY--WHVRLP 62
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 64 AMGMRFVQRYFENSLVRGSKQWQALADARRHGMHVAGYSER-AGGSLYMGQAIFGPDG 122
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 63 A-----QQVYALSEKLADSASLSIKQAKQYAMALGVGLIERDNNNNLNTWVVCMPDG 117
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 123 DLIAARRKIKPTHA-ERTVFGEGDGSHLAVHDTAIG-RLGALCCWEH--IQPLSKYAMYA 178
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 118 SL-----QKHKRLHAFEHVICSQD--QYTVFDTPMGILKIMGILICWDNNLVENARATALLG 171
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 179 AD-----EQVHVASWPSFSLYRGMAVALGPEVNTAASQ----- 211
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 172 ADILLAPHQTGTHSRSPHSMKPIPMALMEN-----RQQDPQALQAAFGGEHKGWL 223
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 212 -----IYAVEGGCYVLASCATVSPEMIKVLVDTPDKEMFLKAGGFGAMIFGPDGRALAE 266
   |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 224 MRWLPARAHNDGMFIFRSNG-----VGRDEEV---RTGNAMVIDPYGRIVKES 269
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 267 LPETEGLVADIDLGMIALAKAADPAQHYRDPVTRLILDRR 310
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 270 C-AIEDDMVTTDIDTLIPESTGRRLTGR--RPELYQILTTTQ 310

RESULT 8
US-10-105-294B-2
; Sequence 2, Application US/10105294B
; Patent No. 6800464
; GENERAL INFORMATION:
; APPLICANT: DRAUZ, KARLHEINZ
; APPLICANT: MAY, OLIVER
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: SYLDATK, CHRISTOPH
; APPLICANT: ALTENBUCHNER, JOSEF
; APPLICANT: WERNER, MARKUS
; APPLICANT: SIEMANN-HERZBERG, MARTIN
; TITLE OF INVENTION: D-Carbamoylase from Arthrobacter crystallopoietes
; FILE REFERENCE: 220024US0X
; CURRENT APPLICATION NUMBER: US/10/105,294B
; CURRENT FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: DE 101 14 999.9
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 315
; TYPE: PRT
; ORGANISM: ARTIFICIAL SEQUENCE
; FEATURE:
; OTHER INFORMATION: SYNTHETIC DNA
; US-10-105-294B-2

Query Match      8.3%; Score 150; DB 2; Length 315;
Best Local Similarity 23.3%; Pred. No. 2.7e-08;
Matches 78; Conservative 42; Mismatches 107; Indels 108; Gaps 17;

QY 26 VEKAIGLIEQAAKQDVRLIAFPE-----TWIPGYPFWIMLGAPAWGMRVQRYFEN 76
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 25 VARLLALBEAASQGAELVFPPELLTFTFFPRTWFEEDF-----EYFDK 70
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 77 SLVRGSKQWQALADARRHGMHVAGYSE-RAAGSLYMGQAIFGPDGLIAARRKLIK-PT 134
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 71 SM--PNDVAPLPERAKDLGVGFYLGAEELTSDEKRYNTSILVNKGHDIVGKYRKMHLLPG 128
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

QY 135 HA-----ERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQV 183  
||| :||| :||| :  
Db 129 HADNREGLPNQHLEKCYFREGD-LRGVYFDHFHGVQVMCLCNDR----- 171  
QY 184 HVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDPDKE 243  
||| :||| :||| :  
Db 172 ---RWP--EYVRSL-----ALQGAELVVLGVNT--PDFVPGWQEEPHAK 208  
QY 244 MF-----LKAGG-----GFAMIFG-----PDGRALAEPLPETEEGLLV 276  
||| :||| :||| :  
Db 209 MFTHLISLQAGAYQNSVFVAAGKSGFEDGHMIGSASVAAPSGEILAKAAGEGDEVVVV 268  
QY 277 -ADIDIGMIALAKAAADPAGHYSPDVTRLLLDR 310  
||| :||| :||| :  
Db 269 KADIDMGK-PYKESVFDFAAH-RRPDAYGIIAERK 301

RESULT 9  
US-09-621-976-5160  
; Sequence 5160, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 5160  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-5160

Query Match 8.1%; Score 145.5; DB 2; Length 175;  
Best Local Similarity 27.3%; Pred. No. 3.4e-08;  
Matches 47; Conservative 28; Mismatches 70; Indels 27; Gaps 7;  
QY 5 MTKYRGAAVQAAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPA 64  
||| :||| :||| :  
Db 1 MTSFRLLAIQLQISSIKD-NVTRACSFIREATQGAKTIVSLPECNSPY----- 49  
QY 65 WGMRFVQRYFENSLVRGSKQWQALADAARRHGMHVAG-YSERAGSLYMGQALFGPDGD 123  
||| :||| :||| :  
Db 50 -GAKYPPRYAEK-IPG-ESTQKLEVAKECSIYILIGGSIPEEDAGKLYNTCAVFGPDGT 105  
QY 124 LIAARRKL-----KPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWE 166  
||| :||| :||| :  
Db 106 LIAKXKRKIHFLDIDVPGKITFQESKTLSPGDS--FSTFDTPYCRVGLGICYD 155

RESULT 10  
US-09-489-039A-10443  
; Sequence 10443, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 10443  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-10443

Query Match 8.0%; Score 144.5; DB 2; Length 337;  
Best Local Similarity 23.5%; Pred. No. 1.3e-07;  
Matches 84; Conservative 56; Mismatches 128; Indels 89; Gaps 20;  
QY 5 MTKY-----GAAVQAAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWI 58  
||| :||| :||| :  
Db 1 MRKYMMEKCI RAATVQFCHRADDKQYNLSVIENFIQAHRENVQLLVPEMCITGY---- 56  
QY 59 WIGAPAWGMRFVQRYFENSL---VRGSKQWQALADAARRHGMHVAGYSERAGS-LYMG 114  
||| :||| :||| :  
Db 57 -----WHVPNLTTETLNDLAERVEDSPSLSRIVTNVAVYHMAIGVGFIKGSDDL YNS 110  
QY 115 QATFGPDGDLIAARRKLKPTHAERTVGE--GDGSHLAVHDTAIG-RLGALCCWEH--I 168  
||| :||| :||| :  
Db 111 YAVCMFDPGK-IHIHKL---HA-----FEHPNISSGNNFTVFDTPWGVRI GILICWDNNLV 162  
QY 169 QPLSKYAMYAAD-----EQVHVASWPS-----FSLYRGMAAYAL 201  
||| :||| :||| :  
Db 163 ENARITALLGADILVAPHQGTGDSRSPFGMKPIEVSIMENRHNDEKLM AAFRGDS-GR 221  
QY 202 GPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDPDKEMFLKAGGFAMIFGPDGR 261  
||| :||| :||| :  
Db 222 GWLMRWLP SR--AHDNGFILFSNG-----VGLDNGE-----IR TGN-AMIIDPYGR 265  
QY 262 ALAEPLPETEEGLVADIDIGMIALAKAAADPAG---HYSRPDVTRLLLDRPAQR 314  
||| :||| :||| :  
Db 266 IVSEAF-SYEDTIVYAEILDIGLTLMS-----TGRRWIHGRREL YGIMCESQYER 315

RESULT 11  
US-09-252-991A-21992  
; Sequence 21992, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21992  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21992

Query Match 7.5%; Score 135.5; DB 2; Length 302;  
Best Local Similarity 22.9%; Pred. No. 1.1e-06;  
Matches 75; Conservative 38; Mismatches 103; Indels 111; Gaps 15;  
QY 21 DLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPAWGMRFVQRYFENSLVR 80  
||| :||| :||| :  
Db 45 DTAHNLERALAIADCA-ADTELVPFPETHLTGFP-----SEDNIA- 84  
QY 81 GSKQWQALAD-----AARRHGMHVAGYSERAGSLYMGQALFGPDGDLIAR 128  
||| :||| :||| :  
Db 85 -----ALAEPLDPTVSAVQRYAERENVSAIGIAEADAGRYNTTLIAPDGIAL--- 135  
QY 129 RKLKPTH--AERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQVHV 185  
||| :||| :||| :  
Db 136 -KYRKTHLWASDRGI FTTPGRNATATLWNGI--RVGLVCFDIEFPESARALGQGAELI I 192  
QY 186 ASWPSFSLYRGMAAYALGPEVNTA-----ASQIYAV-----EGGCYVLASCATVSP E 231  
||| :||| :||| :  
Db 193 VTNGNNDPY-----GPTHRTAIMARAMENQAYAVMVNRVGHGDG----- 232  
QY 232 MIKVLVDTPDKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDIGMIALAKAAA 291





Db 275 IPKSKLVAEILPIHVKKPEQTVVNKNPNVFPSEDDVDQDLFDRGDFAPLK 325

RESULT 15

US-09-252-991A-24644

; Sequence 24644, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24644

; LENGTH: 349

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24644

Query Match

6.5%; Score 117.5; DB 2; Length 349;

Best Local Similarity 25.8%; Pred. No. 0.00017;

Matches 82; Conservative 22; Mismatches 99; Indels 115; Gaps 16;

```

QY 21 DLDRTVEKAIGLIEQAKQDVRLIAFPETWIPGYPFWIWLGAAPAWGMRVQRYFENSLVR 80
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 80 DVTANLAAARRLLEQAAEGCARLAVLPENF-----AAMGRDLAE-LGRAEAR 126

QY 81 GS--KQWQALADAARRHGMHVAVAGYSERAGGSLYMGQAI FGPDGLIA---ARRKLKPT 134
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 127 GNGPILPW--LNSAARDLRLMIVA-----GTLPLPPDGQPEAKANACSLIDE 172

QY 135 HAERTV-----FGEED---GSHLAVHDTAIGRLGALCCWEHIQPLS 172
   | | | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 173 HGERVARYDKLHLPFDVDVADARGRYRESDDYAFGQKITVADTPVGRGLTVTCYDLRFP-- 230

QY 173 KYAMYAADEQVHVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGG----- 218
   | | | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 231 -----ELYTAMREA-GAELITAPSAFTAVTGAAMQVLYRARALET 270

QY 219 -CYVLAS-CATVSPPEMIKVLVDTPKEMFLKAGGCFAMIFGPDGRALAEPLPETEGLLV 276
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 271 QCYLLAAGGGGVHR-----GRETTF-----GHSALVDPWGRVLAE-RPQGEAVLLA 315

QY 277 -----ADIDLGMITALA 287
   | | | | | : | | | : | | | : | | | : | | | : | | | : | | |
Db 316 VRDAAEQADIRRRMPVVA 333

```

Search completed: April 27, 2006, 01:02:21

Job time : 35.448 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:01:20 ; Search time 111.956 Seconds  
(without alignments)  
1291.301 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806  
Sequence: 1 MSEPMTKYRGAIVQAAPVFL.....EDKGDAPALRVVAESAAQAQ 346

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1806	100.0	346	3 US-09-751-299-2	Sequence 2, Appl1
2	1806	100.0	346	4 US-10-146-772-384	Sequence 384, App
3	1806	100.0	346	4 US-10-241-742-384	Sequence 384, App
4	1806	100.0	346	4 US-10-440-523-384	Sequence 384, App
5	1806	100.0	346	4 US-10-440-503-384	Sequence 384, App
6	1806	100.0	346	4 US-10-461-925-384	Sequence 384, App
7	1464	81.1	341	4 US-10-146-772-34	Sequence 34, Appl
8	1464	81.1	341	4 US-10-241-742-34	Sequence 34, Appl
9	1464	81.1	341	4 US-10-440-523-34	Sequence 34, Appl
10	1464	81.1	341	4 US-10-440-503-34	Sequence 34, Appl
11	1464	81.1	341	4 US-10-461-925-34	Sequence 34, Appl
12	1209	66.9	348	4 US-10-146-772-16	Sequence 16, Appl
13	1209	66.9	348	4 US-10-241-742-16	Sequence 16, Appl
14	1209	66.9	348	4 US-10-440-523-16	Sequence 16, Appl
15	1209	66.9	348	4 US-10-440-503-16	Sequence 16, Appl
16	1209	66.9	348	4 US-10-461-925-16	Sequence 16, Appl
17	1165	64.5	353	4 US-10-146-772-212	Sequence 212, App
18	1165	64.5	353	4 US-10-241-742-212	Sequence 212, App
19	1165	64.5	353	4 US-10-440-523-212	Sequence 212, App
20	1165	64.5	353	4 US-10-440-503-212	Sequence 212, App
21	1165	64.5	353	4 US-10-461-925-212	Sequence 212, App
22	1155	64.0	353	4 US-10-146-772-232	Sequence 232, App
23	1155	64.0	353	4 US-10-241-742-232	Sequence 232, App
24	1155	64.0	353	4 US-10-440-523-232	Sequence 232, App
25	1155	64.0	353	4 US-10-440-503-232	Sequence 232, App
26	1155	64.0	353	4 US-10-461-925-232	Sequence 232, App
27	1148	63.6	333	4 US-10-146-772-350	Sequence 350, App

28	1148	63.6	333	4 US-10-241-742-350	Sequence 350, App
29	1148	63.6	333	4 US-10-440-523-350	Sequence 350, App
30	1148	63.6	333	4 US-10-440-503-350	Sequence 350, App
31	1148	63.6	333	4 US-10-461-925-350	Sequence 350, App
32	1132.5	62.7	358	4 US-10-146-772-170	Sequence 170, App
33	1132.5	62.7	358	4 US-10-241-742-170	Sequence 170, App
34	1132.5	62.7	358	4 US-10-440-523-170	Sequence 170, App
35	1132.5	62.7	358	4 US-10-440-503-170	Sequence 170, App
36	1132.5	62.7	358	4 US-10-461-925-170	Sequence 170, App
37	1130	62.6	354	4 US-10-146-772-102	Sequence 102, App
38	1130	62.6	354	4 US-10-241-742-102	Sequence 102, App
39	1130	62.6	354	4 US-10-440-523-102	Sequence 102, App
40	1130	62.6	354	4 US-10-440-503-102	Sequence 102, App
41	1130	62.6	354	4 US-10-461-925-102	Sequence 102, App
42	1125	62.3	345	4 US-10-146-772-334	Sequence 334, App
43	1125	62.3	345	4 US-10-241-742-334	Sequence 334, App
44	1125	62.3	345	4 US-10-440-523-334	Sequence 334, App
45	1125	62.3	345	4 US-10-440-503-334	Sequence 334, App

ALIGNMENTS

RESULT 1

US-09-751-299-2

/ Sequence 2, Application US/09751299

/ Patent No. US20020012974A1

/ GENERAL INFORMATION:

/ APPLICANT: Madden, Mark

/ APPLICANT: Weiner, David P.

/ APPLICANT: Chaplin, Jennifer A.

/ TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE

/ TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS

/ FILE REFERENCE: DIVER1440-2

/ CURRENT APPLICATION NUMBER: US/09/751,299

/ CURRENT FILING DATE: 2000-12-29

/ PRIOR APPLICATION NUMBER: 60/254,414

/ PRIOR FILING DATE: 2000-12-07

/ PRIOR APPLICATION NUMBER: 60/173,609

/ PRIOR FILING DATE: 1999-12-29

/ NUMBER OF SEQ ID NOS: 4

/ SOFTWARE: PatentIn Ver. 2.1

/ SEQ ID NO 2

/ LENGTH: 346

/ TYPE: PRT

/ ORGANISM: Unknown Organism

/ FEATURE:

/ OTHER INFORMATION: Description of Unknown Organism: Obtained from an

/ OTHER INFORMATION: environmental sample

US-09-751-299-2

Query Match	100.0%;	Score 1806;	DB 3;	Length 346;
Best Local Similarity	100.0%;	Pred. No. 2.4e-170;		
Matches 346;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MSEPMTKYRGAIVQAAPVFLDLDRTVEKAIGLIEQAQDVRLIAFPETWIPGYFWIWL	60	
DB	1	MSEPMTKYRGAIVQAAPVFLDLDRTVEKAIGLIEQAQDVRLIAFPETWIPGYFWIWL	60	
QY	61	GAPAWMRFVQRYFENSLVRGSKQWALADARRHGMHVAVGYSERAGSGLYMGQAFGP	120	
DB	61	GAPAWMRFVQRYFENSLVRGSKQWALADARRHGMHVAVGYSERAGSGLYMGQAFGP	120	
QY	121	DGDLTAARRKLKPTHAERTVFGEGDGSILA VHTAIGRLGALCWEHIQPLSKYAMYAAD	180	
DB	121	DGDLTAARRKLKPTHAERTVFGEGDGSILA VHTAIGRLGALCWEHIQPLSKYAMYAAD	180	
QY	181	EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYTASCATVSPMIKVLVDTP	240	
DB	181	EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYTASCATVSPMIKVLVDTP	240	
QY	241	DKEMFLKAGGGFAMIFGPDGRAIAEPLPETEGLLVADIDIGMIALAKAADPAGHSRP	300	
DB	241	DKEMFLKAGGGFAMIFGPDGRAIAEPLPETEGLLVADIDIGMIALAKAADPAGHSRP	300	



Db 241 DKEMFLKAGGGFAMI FGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHSRP 300

QY 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

Db 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

RESULT 2

US-10-146-772-384

; Sequence 384, Application US/10146772

; Publication No. US20030124698A1

; GENERAL INFORMATION:

; APPLICANT: Short, Jay

; APPLICANT: Weiner, David

; APPLICANT: Chaplin, Jennifer

; APPLICANT: Chi, Ellen

; APPLICANT: Milan, Aileen

; APPLICANT: Desantis, Grace

; APPLICANT: Madden, Mark

; APPLICANT: Burk, Mark

; TITLE OF INVENTION: Nitrilases

; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US

; CURRENT APPLICATION NUMBER: US/10/146,772

; CURRENT FILING DATE: 2002-05-15

; PRIOR APPLICATION NUMBER: US 60/309,006

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US 60/351,336

; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 09/751,299

; PRIOR FILING DATE: 2000-12-28

; PRIOR APPLICATION NUMBER: US 60/254,414

; PRIOR FILING DATE: 2000-12-07

; PRIOR APPLICATION NUMBER: US 60/173,609

; PRIOR FILING DATE: 1999-12-29

; NUMBER OF SEQ ID NOS: 386

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 384

; LENGTH: 346

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Obtained from an environmental sample

US-10-146-772-384

Query Match 100.0%; Score 1806; DB 4; Length 346;

Best Local Similarity 100.0%; Pred. No. 2.4e-170;

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLLAFPETWIPGYPFWITL 60

Db 1 MSEPMTKYRGAAVQAAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLLAFPETWIPGYPFWITL 60

QY 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120

Db 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120

QY 121 DGDLLAARRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

Db 121 DGDLLAARRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

Db 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHSRP 300

Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHSRP 300

QY 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

Db 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

RESULT 3

US-10-241-742-384

; Sequence 384, Application US/10241742

; Publication No. US20040002147A1

; GENERAL INFORMATION:

; APPLICANT: Short, Jay

; APPLICANT: Weiner, David

; APPLICANT: Chaplin, Jennifer

; APPLICANT: Chi, Ellen

; APPLICANT: Milan, Aileen

; APPLICANT: Desantis, Grace

; APPLICANT: Madden, Mark

; APPLICANT: Burk, Mark

; TITLE OF INVENTION: Nitrilases

; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US

; CURRENT APPLICATION NUMBER: US/10/241,742

; CURRENT FILING DATE: 2002-09-09

; PRIOR APPLICATION NUMBER: US/10/146,772

; PRIOR FILING DATE: 2002-05-15

; PRIOR APPLICATION NUMBER: US 60/309,006

; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US 60/351,336

; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189

; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 09/751,299

; PRIOR FILING DATE: 2000-12-28

; PRIOR APPLICATION NUMBER: US 60/254,414

; PRIOR FILING DATE: 2000-12-07

; PRIOR APPLICATION NUMBER: US 60/173,609

; PRIOR FILING DATE: 1999-12-29

; NUMBER OF SEQ ID NOS: 386

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 384

; LENGTH: 346

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Obtained from an environmental sample

US-10-241-742-384

Query Match 100.0%; Score 1806; DB 4; Length 346;

Best Local Similarity 100.0%; Pred. No. 2.4e-170;

Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLLAFPETWIPGYPFWITL 60

Db 1 MSEPMTKYRGAAVQAAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLLAFPETWIPGYPFWITL 60

QY 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120

Db 61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120

QY 121 DGDLLAARRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

Db 121 DGDLLAARRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

Db 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP 240

QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHSRP 300

Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAADPAGHSRP 300

QY 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

Db 301 DVTRLLDRRPAQRVVTLLDAFAFEPQNEKGDAPALRVVAESAAAAQ 346

RESULT 4

```
US-10-440-523-384
; Sequence 384, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-440-523-384

Query Match          100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYFPIWTL 60
Db      1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYFPIWTL 60
QY      61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
Db      61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
QY      121 DGDLLIARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
Db      121 DGDLLIARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY      181 EGVHVASWSPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLDTP 240
Db      181 EGVHVASWSPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLDTP 240
QY      241 DKEMFLKAGGFGFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAAADPAGHSRP 300
Db      241 DKEMFLKAGGFGFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAAADPAGHSRP 300
QY      301 DVTIRLLDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAAQ 346
Db      301 DVTIRLLDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAAQ 346

RESULT 5
US-10-440-503-384
; Sequence 384, Application US/10440503
; Publication No. US20040038419A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-440-503-384

Query Match          100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYFPIWTL 60
Db      1 MSEPMTKYRGAAVQAAVFLDLDRTEVEKAIGLIEQAAKQDVRLIAFPETWIPGYFPIWTL 60
QY      61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
Db      61 GAPAWGMRFFVQRYFENSLVRGSKQWQALADAARRHGMHVAVAGYSERAGGSLYMGQAIFGP 120
QY      121 DGDLLIARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
Db      121 DGDLLIARRKCLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY      181 EGVHVASWSPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLDTP 240
Db      181 EGVHVASWSPSFLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLDTP 240
QY      241 DKEMFLKAGGFGFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAAADPAGHSRP 300
Db      241 DKEMFLKAGGFGFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAAADPAGHSRP 300
QY      301 DVTIRLLDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAAQ 346
Db      301 DVTIRLLDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAAQ 346

RESULT 6
US-10-461-925-384
; Sequence 384, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-384

Query Match          100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAFPETWIPGYPFWIML 60
   |||
Db 1 MSEPMTKYRGAAVQAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAFPETWIPGYPFWIML 60

QY 61 GAPAWGMRFFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSGLYMGQAIFGP 120
   |||
Db 61 GAPAWGMRFFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSGLYMGQAIFGP 120

QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
   |||
Db 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTF 240
   |||
Db 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTF 240

QY 241 DKEMFLKAGGFFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRP 300
   |||
Db 241 DKEMFLKAGGFFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRP 300

QY 301 DVTRLLLDRRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346
   |||
Db 301 DVTRLLLDRRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAAQ 346

RESULT 7
US-10-146-772-34
; Sequence 34, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
```

```
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-34

Query Match          81.1%; Score 1464; DB 4; Length 341;
Best Local Similarity 81.7%; Pred. No. 2.1e-136;
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSEPMTKYRGAAVQAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAFPETWIPGYPFWIML 60
   |||
Db 1 MLSPVTQYRAAAVQAAPSFLDLDRTEKTAIIEQAAEQDVRLIAFPETWIPGYPLWIML 60

QY 61 GAPAWGMRFFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSGLYMGQAIFGP 120
   |||
Db 61 GSPAWGMRFFVQRYFENSLVRGSKQWALADARRHGMHVAGYSERAGSGLYMGQAIFGP 120

QY 121 DGDLLAARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
   |||
Db 121 EGELIARRKLPKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTF 240
   |||
Db 181 EQVHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTF 240

QY 241 DKEMFLKAGGFFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRP 300
   |||
Db 241 DKEMFLKAGGFFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRP 300

QY 301 DVTRLLLDRRPAQRVVTLDAAFEPOVEDKGDAP 333
   |||
Db 301 DVTRLLLDRRPAQRVVTLDAAFEPOVEDKGDAP 333

RESULT 8
US-10-241-742-34
; Sequence 34, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-34
```





QY	181	E QVHVASWPFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKYLVDTP	240
		:    :    :    :	
Dd	181	E QVHVASWPCFSLYRGMAAYALGPEVNTAASQYYAVEGGCYVLASCLVVTPEILKYLIDTP	240
QY	241	DKEMFLKAGGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHSRP	300
		:    :    :    :	
Dd	241	DKEPLLLAGGGFSMIFGPDGRALAQPLPETEBGLVTAETDGLATLAKAAADPAGHTARP	300
QY	301	DVTRLLLDRRPAQRVVTLDAFEPOÑEDKKDAP	333
		:    :    :	
Dd	301	DVTRLLLNPRPARVEALGPRFEVVQSEQAEP	333

```

RESULT 11
US-10-461-925-34
; Sequence 34, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R) -ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-34

```

Query Match	81.1%;	Score 1464;	DB 4;	Length 341;	
Best Local Similarity	81.7%;	Pred. No. 2.1e-136;			
Matches 272;	Conservative 28;	Mismatches 33;	Indels 0;	Gaps 0;	
QY	1	MSEPMTKYRGAAVQAAAPVFLDLDRTEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIWL	60		
		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :			
Db	1	MLSPVTQYRAAAVQAAAPSFLLDLDRTEKTIATIEQAAEQVRLIAPETWIPGYPLWIWL	60		
QY	61	GAPAWGMRVQRYFENSLVRGSKQWQALADAARRHGMHVAVGYSERAGSLYMGQAIFGP	120		
		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :			
Db	61	GSPAWGMRVQRYFENSLVRGSKQWNAIADAARRHMTTVVGFSEERAGSLYMGQAIFGP	120		
QY	121	DGDLIARRKLKPTHAERTVFEGEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD	180		
		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :			
Db	121	EGELIARRKLKPTHAERTVFEGEDGSHLAVYETGVGRIGALCWEHIQPLSKYAMYAAN	180		
QY	181	EQVHVASWPSFSLYRCMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTP	240		
		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :			
Db	181	EQVHVASWPCFSLYRCMAAYALGPEVNTAASQVYAVEGGCYVLASCLVVTPEILKVLIDTP	240		
QY	241	DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGMTALAKAAADPAGHYSRP	300		
		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :			
Db	241	DKEPLLLAGGGFSMIFGPDGRALAAQPLPETEEGLVTAETIDGAILAKAAADPAGHYARP	300		
QY	301	DVTRLLLDRRPAQRVVTLDDAAFEFPQNEKGDAP	333		
		: : : : :   : : : : :   : : : : :			
Db	301	DVTRLLLNRPAPARVEALGPRFEVVOSEQAEP	333		

RESULT 12  
US-10-146-772-16  
; Sequence 16, Application US/10146772  
; Publication No. US20030124698A1

; GENERAL INFORMATION:  
 ; APPLICANT: Short, Jay  
 ; APPLICANT: weiner, David  
 ; APPLICANT: Chaplin, Jennifer  
 ; APPLICANT: Chi, Ellen  
 ; APPLICANT: Milan, Aileen  
 ; APPLICANT: Desantis, Grace  
 ; APPLICANT: Madden, Mark  
 ; APPLICANT: Burk, Mark  
 ; TITLE OF INVENTION: Nitriases  
 ; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
 ; CURRENT APPLICATION NUMBER: US/10/146,772

```

      / PRIOR APPLICATION NUMBER: US 60/309, 006
      / PRIOR FILING DATE: 2001-07-30
      / PRIOR APPLICATION NUMBER: US 60/351, 336
      / PRIOR FILING DATE: 2002-01-22
      / PRIOR APPLICATION NUMBER: US 60/300, 189
      / PRIOR FILING DATE: 2001-06-21
      / PRIOR APPLICATION NUMBER: US 09/751, 299
      / PRIOR FILING DATE: 2000-12-28
      / PRIOR APPLICATION NUMBER: US 60/254, 414
      / PRIOR FILING DATE: 2000-12-07
      / PRIOR APPLICATION NUMBER: US 60/173, 609
      / PRIOR FILING DATE: 1999-12-29
      / NUMBER OF SEQ ID NOS: 386
      / SOFTWARE: FastSeq for Windows Version 4.0
      / SEQ ID NO 16
      / LENGTH: 348
      / TYPE: PRT
      / ORGANISM: Unknown
      / FEATURE:
      / OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-16

```

Query Match	66.9%;	Score 1209;	DB 4;	Length 348;
Best Local Similarity	67.1%;	Pred. No. 4.5e-11;		
Matches 230;	Conservative 42;	Mismatches 69;	Indels 2;	Gaps 2;

  

QY	4	PMTK-YRGAAVQAPVFLDLDRTEKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGA	62
		: :               : :           :                   : :	
DB	2	PTSKQFRAVAVQAPVFLDLLEGAISKISILIEEASNGAKLIAFPETWIPGYPMWIWLDLS	61
QY	63	PAMGMRFOVORYFENSLVRGSKQWQALADABRHGMHVAVYSERAGSLYMGQAFGPDG	122
		:   : : : : : : : : : : : : : : :	
DB	62	PAMGMRFOVORYFDNSLMLGSEQAKRMNQAAANNKIYVMGYSERSGSLYMGQSIINDKG	121
QY	123	DLIAARRKLKPTHAERTVFEGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQ	182
		:                                 :   :   :   :   :   :   :	
DB	122	ETIFTRRLKPTHERTVFEGEGDGSHLQVMDTEIGRVGAMCCWEHLQPLSKYAMYSQDEQ	181
QY	183	VHVASWPSFSLYRGMAAYALGPENVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPK	242
		: :                         : :   : :   : :   : :   : :	
DB	182	IHIASWPSFSLYRGAAYALGPBLNNAASQMYAAEGCCFVLAPCATVSKEMIEMLIDDPK	241
QY	243	EMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGIMIALAKAAADPAGHYSRPDV	302
		:   :   :   :   :   :   :   :   :   :	
DB	242	EPLLLEGGGFTWICYPPDGRPLAKPLPENEBGLLYADIDLGIMISMAKAAADPAGHYARPDV	301
QY	303	TRLILDRRPAQRVVTLDAAFEPOINEDKGDAPALRVVAESAASAAA	345
		:       : :   : :   : :   : :   : :	
DB	302	TRLLENAPANRVEYINPASGP-TESLKDMGKQMEAEQOQKAA	343

```

RESULT 13
US-10-241-742-16
; Sequence 16, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen

```

```
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-16
```

```
Query Match          66.9%; Score 1209; DB 4; Length 348;
Best Local Similarity 67.1%; Pred. No. 4.5e-111;
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;
```

```
QY      4 PMTK-YRGAAVQAPVFLDLDRTEKAIGLIEQAKQDVRLIAFPETWIPGYPFWIWLG 62
      2 PTSKQFRVAAVQAPVFLDLLEGAISKGISLIEAASNGAKLIAFPETWIPGYPMWIWLD 61
QY      63 PAMGMRVQRYFENSILVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQAI 122
      62 PAMGMRVQRYFDNSLMLGSEQAKRMNQAAANNKIYVMGYSERSGSGLYMGQSIINDKG 121
QY      123 DLIAARRKLKPTHAERTVFGEEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYADEQ 182
      122 ETIFTRRLKPTHVERTVFGEEDGSHLCVMDTEIGRVGAMCWEHLQPLSKYAMYSQDEQ 181
QY      183 VHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPBEMIKVLVDPDK 242
      182 IHIASWPSFSLYRGAAYALGPELNNAAASQMYAAEGQCFVLAPCATVSKEMIEMLIDDPK 241
QY      243 EMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDV 302
      242 EPLLEGGGFTMITYGPDPGRPLAKPLPENEGLLYADIDLGMIISMAKAAADPAGHYARPDV 301
QY      303 TRLLDRRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAA 345
      302 TRLLFNSAPANRVEYINPASGP-TESTLKDGMQMGEAEQOKAA 343
DB
```

```
RESULT 14
US-10-440-523-16
; Sequence 16, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
```

```
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-16
```

```
Query Match          66.9%; Score 1209; DB 4; Length 348;
Best Local Similarity 67.1%; Pred. No. 4.5e-111;
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;
```

```
QY      4 PMTK-YRGAAVQAPVFLDLDRTEKAIGLIEQAKQDVRLIAFPETWIPGYPFWIWLG 62
      2 PTSKQFRVAAVQAPVFLDLLEGAISKGISLIEAASNGAKLIAFPETWIPGYPMWIWLD 61
QY      63 PAMGMRVQRYFENSILVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQAI 122
      62 PAMGMRVQRYFDNSLMLGSEQAKRMNQAAANNKIYVMGYSERSGSGLYMGQSIINDKG 121
QY      123 DLIAARRKLKPTHAERTVFGEEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYADEQ 182
      122 ETIFTRRLKPTHVERTVFGEEDGSHLCVMDTEIGRVGAMCWEHLQPLSKYAMYSQDEQ 181
QY      183 VHVASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPBEMIKVLVDPDK 242
      182 IHIASWPSFSLYRGAAYALGPELNNAAASQMYAAEGQCFVLAPCATVSKEMIEMLIDDPK 241
QY      243 EMFLKAGGGFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRPDV 302
      242 EPLLEGGGFTMITYGPDPGRPLAKPLPENEGLLYADIDLGMIISMAKAAADPAGHYARPDV 301
QY      303 TRLLDRRPAQRVVTLDAAFEPOVEDKGDAPALRVVAESAAAA 345
      302 TRLLFNSAPANRVEYINPASGP-TESTLKDGMQMGEAEQOKAA 343
DB
```

```
RESULT 15
US-10-440-503-16
; Sequence 16, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE
```

; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440,503  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; PRIOR FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16  
; LENGTH: 348  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-16

Query Match 66.9%; Score 1209; DB 4; Length 348;  
Best local Similarity 67.1%; Pred. No. 4.5e-111;  
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

QY	4	PMTK-YRGAAYQAAPVFLDLDRIVEKAIGLIEQAKQDVRLIAFPETWIPGYPFWIWLG	62
DB	2	PTSKQFRVAAYQAAPVFLDLGSAISKISLIEEASNGAKLIAFPETWIPGYPWITWLD	61
QY	63	PAMGMRFYQRYEENSLVRGSKQWQALDAARRHGMHVAVAGYSERAGSLYMGQAI	122
DB	62	PAMGMRFYQRYFDNSLMLGSEQAKRMWQAANNKIYVVMGYSERGGSLYMGQSI	121
QY	123	DLIAARRKLPKPTHAERTVFEGEGDGHIAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQ	182
DB	122	ETIFTRRKLKPTHTVERTVFEGEGDGHLCVMDTEIGRVGAMCCWEHIQPLSKYAMYSQDEQ	181
QY	183	VHVASWPSFSLYRGMAIYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDT	242
DB	182	IHIASWPSFSLYRGAAYALGPENLNAASQMYAAEGQCFVLAPCATVSKEMIEMLI	241
QY	243	EMFLKAGGFFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSR	302
DB	242	EPLLEGGFTWYIGPDGRPLAKPLPENEGLLYADIDLGMISMAKAAADPAGHYAR	301
QY	303	TRLLDRRRPAQRVVTLDAAFEQNEDEKGDAPALRVVAESAAAA	345
DB	302	TRLLFNSAPANRVEYINPASGP-TESLXDMGKMQEAEQOKAA	343

Search completed: April 27, 2006, 01:06:08  
Job time : 113.956 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:02:36 ; Search time 18.2372 Seconds  
(without alignments)  
862.816 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806

Sequence: 1 MSEPMTKYRGAAVQAPVFL.....EDKGDA PALRVVAESA AAAQ 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New: \*  
1: /SIDS5/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /SIDS5/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /SIDS5/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /SIDS5/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
5: /SIDS5/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
6: /SIDS5/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /SIDS5/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /SIDS5/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	869	48.1	356	6	US-10-537-075-7 Sequence 7, Appli
2	810	44.9	369	6	US-10-919-182-16 Sequence 16, Appl
3	809	44.8	369	6	US-10-919-182-18 Sequence 18, Appl
4	807	44.7	369	6	US-10-919-182-4 Sequence 4, Appli
5	807	44.7	369	6	US-10-919-182-6 Sequence 6, Appli
6	805	44.6	369	6	US-10-919-182-12 Sequence 12, Appl
7	804	44.5	369	6	US-10-919-182-14 Sequence 14, Appl
8	796	44.1	369	6	US-10-919-182-8 Sequence 8, Appli
9	430.5	23.8	333	7	US-11-096-568A-20687 Sequence 20687, A
10	430.5	23.8	351	7	US-11-096-568A-20686 Sequence 20686, A
11	399.5	22.1	330	7	US-11-096-568A-6955 Sequence 6955, Ap
12	399.5	22.1	350	7	US-11-096-568A-6954 Sequence 6954, Ap
13	315	17.4	233	7	US-11-096-568A-6956 Sequence 6956, Ap
14	284	15.7	193	7	US-11-096-568A-20688 Sequence 20688, A
15	131.5	7.3	901	7	US-11-087-099-2095 Sequence 2095, Ap
16	131.5	7.3	901	7	US-11-188-298-2049 Sequence 2049, Ap
17	110.5	6.1	304	6	US-10-467-657-7410 Sequence 7410, Ap
18	107.5	6.0	532	7	US-11-079-463-5688 Sequence 5688, Ap
19	106.5	5.9	295	7	US-11-079-463-5457 Sequence 5457, Ap
20	105	5.8	259	7	US-11-045-004-822 Sequence 822, App
21	102.5	5.7	336	7	US-11-188-298-20349 Sequence 20349, A
22	99.5	5.5	339	7	US-11-188-298-16658 Sequence 16658, A
23	95.5	5.3	553	7	US-11-188-298-19842 Sequence 19842, A
24	94	5.2	306	7	US-11-188-298-17485 Sequence 17485, A
25	93	5.1	3655	7	US-11-075-185-5 Sequence 5, Appli

26	91.5	5.1	1632	6	US-10-506-454-1147	Sequence 1147, Ap
27	91	5.0	457	7	US-11-059-867-14	Sequence 14, Appl
28	90	5.0	811	7	US-11-200-2968-81	Sequence 81, Appl
29	89.5	5.0	290	7	US-11-188-298-1550	Sequence 1550, Ap
30	89.5	5.0	461	7	US-11-087-099-4404	Sequence 4404, Ap
31	89	4.9	647	6	US-10-915-002-283	Sequence 283, App
32	89	4.9	1192	6	US-10-858-730-72	Sequence 72, Appl
33	88.5	4.9	356	7	US-11-188-298-22086	Sequence 22086, A
34	88.5	4.9	516	7	US-11-096-568A-14918	Sequence 14918, A
35	88.5	4.9	541	7	US-11-096-568A-14917	Sequence 14917, A
36	88.5	4.9	601	7	US-11-096-568A-14916	Sequence 14916, A
37	88	4.9	436	7	US-11-188-298-476	Sequence 476, App
38	87.5	4.8	347	7	US-11-188-298-1186	Sequence 1186, Ap
39	87.5	4.8	492	7	US-11-188-298-18192	Sequence 18192, A
40	87	4.8	321	7	US-11-188-298-1484	Sequence 1484, Ap
41	87	4.8	627	7	US-11-188-298-6685	Sequence 6685, Ap
42	87	4.8	867	7	US-11-079-463-10013	Sequence 10013, A
43	86.5	4.8	301	7	US-11-072-512-2223	Sequence 2223, Ap
44	86.5	4.8	351	7	US-11-096-568A-21830	Sequence 21830, A
45	86.5	4.8	353	7	US-11-096-568A-21829	Sequence 21829, A

ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesselner, Maria  
; APPLICANT: Zelinski, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-10-537-075-7

QY	9	RGAAVQAPVFLDTRTVEKAIGLIEQAAKQDVRLIAPPETWIPGYPFWIWLGA	PAWGMR	68
DB	8	RAAAVQASPNDLATGVDKTELARQARDEGCDLIVGETWLPGFPHWLGA	PAWSLK	67
QY	69	FWQRYFENSLVRGSKWQALADAARRHGMHVAGYSERAGSLYMGQAFGPDGDLIAAR		128
DB	68	YSARYYANSLSDSAEFORIAQARTLGIFALGYSEKSGSLYLGQCLIDDKGEMLSR		127
QY	129	RKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYADEQVHVASW		188
DB	128	RKLKPTHTVERTVFGEGYARDLIVSDTELGVALCCWEHLSPLSKYALYSQHEAIIHIAW		187
QY	189	PSFSLYRGMAFYALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVTPDKEMFLXA		248
DB	188	PSFSLYSEQAHALSAKVMAASQIYSVEGGCTTIAASSVVTQETLDMLEVEGHNAPLIKV		247
QY	249	GGGFAMIFGPDGRLAEPPLPETEGLLVADIDGMITALAKAADPAGHYRSPDYTRLILD		308
DB	248	GGGSSMIFAPDGRITLAPYLPHDAEGLIADLNMEIAPAKAINDPVGHSKPEATRLVLVD		307



QY 309 ---RRPAQRVVTLDAAFEQPONEDKGADAPALRVVAESAAAA 345  
| | | | | : : : : : | : : : : :  
Db 308 LGHRDPMTRV-----HSKSVTREAPFEGVQSKIASVA 340

RESULT 2  
US-10-919-182-16  
; Sequence 16, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. dupont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACIC USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 16  
; LENGTH: 369  
; TYPE: PRF  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change  
US-10-919-182-16

Query Match 44.9%; Score 810; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 3.9e-66;  
Matches 160; Conservative 53; Mismatches 126; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPAW 65  
: : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 6 SKFLAATVQAEPVWLADADATIDKSIGIIEEAAQKASLIAFPEVFIPGYPWAMLGDVKY 65  
QY 66 GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQAIFGPDGDLI 125  
: | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 66 SLSTSRHYHENSLELGDDRMRRLQLAARRNKIALVMGYSERAGSRYLSQVFI DERGEIV 125  
QY 126 AARRKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADQVHV 185  
| | | | | | | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 126 ANRRKLKPTHTVERTIYGEENGTDPLTHDFAFGRVGLNCWEHVQPLSKFMYMSLGEQVHV 185  
QY 186 ASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPPDKEMF 245  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 186 ASWPAMSPLQPDVFQLSIEANATVTRSYAIEGQTFVLCSTQVIGPSAIEFTCLNDEQRAL 245  
QY 246 LKAGGPFAMIFGPDGRALAEPLPETEEGLVADIDLGMIALAKAADPAGHYSRPDVTRL 305  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 246 LPQCGWARIYGPDSSELAKPLAEDAEGLIYAEIDLQIILAKAGADPVGHYSRPDVLSV 305  
QY 306 LLDR---PAQRVVTLDAAFEQPONEDKGADAPALRVVAESAAAA 345  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 306 QFDPNRNHTPVHR-IGIDGRLDVNTRSRVENFRLRQAAEQERQA 347

RESULT 3  
US-10-919-182-18  
; Sequence 18, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. dupont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACIC USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 18

; LENGTH: 369  
; TYPE: PRF  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
US-10-919-182-18

Query Match 44.8%; Score 809; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 4.8e-66;  
Matches 160; Conservative 53; Mismatches 126; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPAW 65  
: : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 6 SKFLAATVQAEPVWLADADATIDKSIGIIEEAAQKASLIAFPEVFIPGYPWAMLGDVKY 65  
QY 66 GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQAIFGPDGDLI 125  
: | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 66 SLSTSRHYHENSLELGDDRMRRLQLAARRNKIALVMGYSERAGSRYLSQVFI DERGEIV 125  
QY 126 AARRKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADQVHV 185  
| | | | | | | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 126 ANRRKLKPTHTVERTIYGEENGTDPLTHDFAFGRVGLNCWEHVQPLSKFMYMSLGEQVHV 185  
QY 186 ASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVDTPPDKEMF 245  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 186 ASWPAMSPLQPDVFQLSIEANATVTRSYAIEGQTFVLCSTQVIGPSAIEFTCLNDEQRAL 245  
QY 246 LKAGGPFAMIFGPDGRALAEPLPETEEGLVADIDLGMIALAKAADPAGHYSRPDVTRL 305  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 246 LPQCGWARIYGPDSSELAKPLAEDAEGLIYAEIDLQIILAKAGADPVGHYSRPDVLSV 305  
QY 306 LLDR---PAQRVVTLDAAFEQPONEDKGADAPALRVVAESAAAA 345  
| | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 306 QFDPNRNHTPVHR-IGIDGRLDVNTRSRVENFRLRQAAEQERQA 347

RESULT 4  
US-10-919-182-4  
; Sequence 4, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. dupont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACIC USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 4  
; LENGTH: 369  
; TYPE: PRF  
; ORGANISM: Acidovorax facilis 72w  
US-10-919-182-4

Query Match 44.7%; Score 807; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 7.3e-66;  
Matches 160; Conservative 52; Mismatches 127; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAAPVFLDLDRTEVEKAIGLIEQAAKQDVRLIAPETWIPGYPFWIMLGAPAW 65  
: : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 6 SKFLAATVQAEPVWLADADATIDKSIGIIEEAAQKASLIAFPEVFIPGYPWAMLGDVKY 65  
QY 66 GMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSGLYMGQAIFGPDGDLI 125  
: | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 66 SLSTSRHYHENSLELGDDRMRRLQLAARRNKIALVMGYSERAGSRYLSQVFI DERGEIV 125  
QY 126 AARRKLKPTHAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAADQVHV 185  
| | | | | | | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | | :  
Db 126 ANRRKLKPTHTVERTIYGEENGTDPLTHDFAFGRVGLNCWEHVQPLSKFMYMSLGEQVHV 185

[illegible]

```

RESULT 5
US-10-919-182-6
; Sequence 6, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B2 and H9
US-10-919-182-6

```

[illegible]

RESULT 6  
US-10-919-182-12  
; Sequence 12, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark

```

; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
US-10-919-182-12

```

Query Match	44.6%;	Score 805;	DB 6;	Length 369;
Best Local Similarity	46.6%;	Pred. No. 1.1e-65;		
Matches 160;	Conservative 51;	Mismatches 128;	Indels 4;	Gaps 2;
QY	6	TKYRGAAVQAAPFVFLDLDRIVEKAIGLIEQAAKQDVRLIAFPETWIPGYPFMTLGAAPAW	65	
	:		:	
Db	6	SKFLAATVQAEPPWMLDADATITDKSIGLIEEAAQKASLIAFPVEVFIIPGYPMWMLGDVKY	65	
QY	66	GMRFYORYFENS LVRGSKQWQALADAAARRHGHHVAVGYSERAGSLYMGQALFEGDDLI	125	
	:		:	
Db	66	SLSFYSRYHENSLELGDDRMRLQLAARRNKIALVMGYSEREAGSRYLSQVFIDERGEIV	125	
QY	126	AARRKLTPTHABRTVFEGEGDGSHLAVHDTAIGRLGALCCWEHIOPLSKYAMYAADQVHV	185	
	:		:	
Db	126	ANRRRLKPTHVERTIYGEGNGIDFLTHDPAFGRVGLNCWEHMFQPLSKFMWYSLGEQVHV	185	
QY	186	ASWPSFSLYRGMAAYALGPEVNTAASQIYAVEGGCYVLASCATVSPEMIKVLVTPDKEMF	245	
	:		:	
Db	186	ASWPAMSPLOPDPVFLQSLIEANATVCRSYAIEGQTFVLCSTQVITGPSAIEFTCLNDEQRAL	245	
QY	246	LKAGGCFAMIFGPDGRALAEPLPETEEGLLVADIDLGMIALAKAAADPAGHYSRBDVTRL	305	
	:		:	
Db	246	LPQGGCMARITYGPDGSELAKPLAEDAEGILYAEIDLEQIILLAKAGADPVGHYSRBDVLSV	305	
QY	306	LLDR---PAQRVVTLLDAAFEPQNEKGDAPALRVVAESAATA	345	
	:		:	
Db	306	QFDPNRNHTPVHR-IGIDGRLLDVNTRSRVENFRRLRQAALAEQERQA	347	

```

RESULT 7
US-10-919-182-14
; Sequence 14, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
US-10-919-182-14

```

Query Match	44.5%;	Score 804;	DB 6;	length 369;
Best Local Similarity	46.6%;	Pred. No. 1,4e-65;		
Matches 160;	Conservative 52;	Mismatches 127;	Indels 4;	Gaps 2;
QY	6	TKYRGAAVQAPVFLDLDRTEVKAIGLIEQAAKQDVRLLIAPETWIPGYPFMIWLGAPAW	65	
	:	:	:	:

Db 6 SKFLAATVQAEFPWLADATIDKISIGIIEEAAQKASLIAPEVEFIGPYWAWLGDV KY 65

QY 66 GMRFYQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGSLYMGQAI FGPDDLI 125

Db 66 SLSTSRHYENSLELGDDRMRLQLAARNKIALVMGYSERAGSRYLSQV FIDERGEIV 125

QY 126 AARRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIOPLSKYAMYADEQVHV 185

Db 126 ANRRKLKPTHAERTIYGEENGDTFLTHDFAFGRVGLNCWEHKQPLSKFMYSLGEQVHV 185

QY 186 ASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTPKEMF 245

Db 186 ASWPAMSPLOPDVFOLSIEANATVTRSYAIEGQTFVLCSTQVIGPSAIEFTCLNDEQRAL 245

QY 246 LKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADPAGHYSRPDVT RL 305

Db 246 LPQCGWARIYGPDSGLAKPLAEDAEGILYAEIDLQILLAKAGADPVGHYSRPDVL SV 305

QY 306 LDDR---PAQRVVTLDAAFEPOBEDKGDAPALRVVAESAAAA 345

Db 306 QFDBRNHTPVHR-IGIDGRLDVNTRSRVENFRLRQAAEQERQA 347

RESULT 8

US-10-919-182-8

; Sequence 8, Application US/10919182

; Publication No. US20060035352A1

; GENERAL INFORMATION:

; APPLICANT: E.I. duPont de Nemours and Company, Inc.

; APPLICANT: Di Cosimo, Robert

; APPLICANT: Payne, Mark

; APPLICANT: O'Keefe, Daniel

; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACIC USING NITRILASE MUTANTS

; FILE REFERENCE: CL2584 US NA

; CURRENT APPLICATION NUMBER: US/10/919,182

; CURRENT FILING DATE: 2004-08-16

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: Patent in version 3.2

; SEQ ID NO 8

; LENGTH: 369

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Mutant nitrilase B4

US-10-919-182-8

Query Match 44.1%; Score 796; DB 6; Length 369;

Best Local Similarity 46.6%; Pred. No. 7.4e-65;

Matches 160; Conservative 49; Mismatches 130; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAPVFLDLDRTEVEKAIGLIEQAKQDVRLIAFPETWIPGYFWIMLGAPAW 65

Db 6 SKFLAATVQAEFPWLADATIDKISIGIIEEAAQKASLIAPEVEFIGPYWAWLGDV KY 65

QY 66 GMRFYQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGSLYMGQAI FGPDDLI 125

Db 66 SLSTSRHYENSLELGDDRMRLQLAARNKIALVMGYSERAGSRYLSQV FIDERGEIV 125

QY 126 AARRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIOPLSKYAMYADEQVHV 185

Db 126 ANRRKLKPTHAERTIYGEENGDTFLTHDFAFGRVGLNCWEHKQPLSKIMMYSLGEQVHV 185

QY 186 ASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTPKEMF 245

Db 186 ASWPAMSPLOPDVFOLSIEANATVTRSYAIEGQTFVLCSTQVIGPSAIEFTCLNDEQRAL 245

QY 246 LKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADPAGHYSRPDVT RL 305

Db 246 LPQCGWARIYGPDSGLAKPLAEDAEGILYAEIDLQILLAKAGADPVGHYSRPDVL SV 305

QY 306 LDDR---PAQRVVTLDAAFEPOBEDKGDAPALRVVAESAAAA 345

Db 306 QFDBRNHTPVHR-IGIDGRLDVNTRSRVENFRLRQAAEQERQA 347

RESULT 9

US-11-096-568A-20687

; Sequence 20687, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; TITLE OF INVENTION: Therby

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 20687

; LENGTH: 333

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(333)

; OTHER INFORMATION: Ceres Seq. ID no. 12386686

US-11-096-568A-20687

Query Match 23.8%; Score 430.5; DB 7; Length 333;

Best Local Similarity 35.0%; Pred. No. 1.5e-31;

Matches 117; Conservative 52; Mismatches 128; Indels 37; Gaps 8;

QY 6 TKYRGAAVQAPVFLDLDRTEVEKAIGLIEQAKQDVRLIAFPETWIPGYP---FWIMLG 61

Db 10 TTARVTVQASSVFYDTPATLDKAEKLVAAEAGYSQVLFPFVFGVPHGSTFGILVVG 69

QY 62 -APAWGMRFYQRYFENSLVRGSKQWQALADARRHGMHVAVAGYSERAGSLYMGQAI FGP 120

Db 70 NRTAKGKEDRQKYHASAIDVPGEVSRLSALAGKYKFLVIGVERAGYTLVNTVLSFDP 129

QY 121 DGDLLAARRKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCCWEHIOPLSKYAMYAAD 180

Db 130 LGKYLGHKRKRWPTALERVFWGFGDGTTPVDTPIGKMGALICWENRMPILRTAMYAKG 189

QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQI-----YAVEGGCYVLASCA-----TV 228

Db 190 IBIYCA-----PTVDCMPTWLSMTHIALEGGCFVLSACQFCRKNYP 233

QY 229 SPEMIKV-LVDTPKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALA 287

Db 234 PPEYTFGLBEESPESVVCSSG--SVIISPLGTVLGAPNYES-EALITADLDLGEIVRA 290

QY 288 KAADPAGHYSRPDVT RLLDRPAQRVVTLDAA 321

Db 291 KFDPDVGHYSRPEVLSLVKSDPKPAVSFISAA 324

RESULT 10

US-11-096-568A-20686

; Sequence 20686, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; TITLE OF INVENTION: Therby

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 20686

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(351)

; OTHER INFORMATION: Ceres Seq. ID no. 12386685





```

; LENGTH: 233
; TYPE: prt
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(233)
; OTHER INFORMATION: Ceres Seq. ID no. 15169319
;
US-11-096-568A-6956

```

Query Match	17.4%	Score 315;	DB 7;	Length 233;
Best Local Similarity	35.5%;	Pred. No. 3.3e-21;		
Matches	88;	Conservative	40;	Mismatches 94;
				Indels 26;
				Gaps 9;

[illegible]

```

RESULT 14
US-11-096-568A-20688
; Sequence 20688, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20688
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(193)
; OTHER INFORMATION: Ceres Seq. ID no. 12386687
US-11-096-568A-20688

```

Query Match	15.7%;	Score 284;	DB 7;	Length 193;
Best Local Similarity	37.1%;	Pred. No. 1.8e-18;		
Matches	75;	Conservative	28;	Mismatches 67;
			Indels	32;
			Gaps	6;

```
QY      133 PTHAERTVFEGEDGSHLAVDTAIGRLGALCCWEHIQPLSKYAMYAADCEQHVASWPFSF 192
       || | : | | | | : | | | | : | | | | : | | | | : | | |
Db      2 PTALERVFWFGFDGSTIPVDTPIGKMGAICWENRMPLLRTAMYAKGIEIYCA----- 55

QY      193 LYRGMAYALGPVENTAAOI-----YAVEGGCYVLASCA-----TVSPEMIKV-LVDT 239
       | | | : | : | : | : | | | | : | | | | : | | |
Db      56 -----PTVDCMPTWLSSMTHTIALEGCFVLISACQCRRKNYP PPPEYTFCGLEEE 105

QY      240 PDKEMFLKAGGGFAMI FPGDGRALAEP LPETEBGLLVADIDLGMIATAKAADPACHYSR 299
       | | : : | | : | | | | : | | | | : | | | | : | | | |
Db      106 PSPESVVCSG--SVIIISP LGTIVLAGPNYES-EALLTADI DLGEIVRAKFDFDVGHISR 162

QY      300 PDVT RLLDRPPAQRV VTLDA A 321
```

Db 163 PEVLSLVKSDPKPAVSFISAA 184

```

RESULT 15
US-11-087-099-2095
; Sequence 2095, Application US/11087099
; Publication No. US20060041961A1
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S. et al.
; TITLE OF INVENTION: Genes and Uses for Plant Improvement
; FILE REFERENCE: 38-21(53450) B EP
; CURRENT APPLICATION NUMBER: US/11/087,099
; CURRENT FILING DATE: 2005-03-22
; NUMBER OF SEQ ID NOS: 12464
; SEQ ID NO 2095
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Xanthomonas axonopodis pv. citri str. 306
US-11-087-099-2095

```

Query Match	7.3%;	Score 131.5;	DB 7;	length 901;
Best Local Similarity	22.9%;	Pred. No. 0.0011;		
Matches	86;	Conservative	37;	Mismatches 109;
			Indels	143;
			Gaps	17;

QY	2	SEPMTKYYR-----	GAAVQAAPVFLDLDRIVE- KAIGLI-----	EQAAKQD	40
		: :	: :     : :	:     :	
Dd	3	SHPTFRFRSLAQGVLLALCATGAHAHAATPPYLDTORQSFEEARLAADLVSRMTLEEKAQAQMQ			62
QY	41	VRLIAFPETWIPGYPFWIMLGAPAWGMRFVQRYFFENSLVRGSKQWQALADARRHGMHV			100
		: :     :		:	
Dd	63	NAAPALPRLQVPAYDWW-----		NEALHGV	86
QY	101	AGYSERAGSGLYMGQALFGPDGDLIARRKLIKPTHAERTVFEGEDGSHLAVHDTAIGRLG			160
		:	:	:     :	
Dd	87	A-----RAGGATVFPQAIG-----	MAATFDLPLMHEVAT-----	AISDEARAK--	124
QY	161	ALCCMEHIQPL--SKYAMYAADEQVHAVSW-PSFSLYRGMAVALGPEVNTAASQIYAVEG			217
		: :	:     : :	:	
Dd	125	-----HHQFLRQNHARYQG-----	LTFWSPNINI FRDPRWGRGQE-----		160
QY	218	GCYVLASCATVSPEMIKVLVDTPDKEMFLKAGGGFAMI FGPDGRALAEPLPETEGLLVA			277
		: :	:     :	:	
Dd	161	-----TYGEDPFLTARMGVTFVQGLQEGGAA--	APKNAQGEPPYR		197
QY	278	DIDLGMIALAKAADPAGHYSRPDVTRLLDRRPAQRVV--	TLDAAFEP-QNEKDGA--		332
		:         :	:		
Dd	198	KLD-----	ATAKHLAVHSGPEADRHHPDARPSQRLYETYYLPAFEALVKEGKYDAVM		243
QY	333	-PALRVVAESAAAAQ	346		
		: :			
Dd	250	GAYNRVYGESASASK	264		

Search completed: April 27, 2006, 01:06:50  
Job time : 20.2372 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:18:51 ; Search time 7.7962 Seconds  
(without alignments)  
2502.854 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1826  
Sequence: 1 atgaaagaagctatcaaggt.....cgcaactcgaagaatttga 1014

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090036\_18677/app\_query.fasta\_1  
-DB=PIR -QFMT=fastan -SUFFIX=trp -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US09751299 @CGN\_1\_1\_77 @runat\_26042006\_090036\_18677 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	836.5	45.8	366	2	A45070	nitriase - Rhodoc
2	819	44.9	356	2	A47181	nitriase (EC 3.5.
3	800	43.8	383	2	A43470	aliphatic nitriase
4	638	34.9	354	2	UC4212	nitriase (EC 3.5.
5	634.5	34.7	349	1	A28658	nitriase (EC 3.5.
6	559	30.6	368	2	UQ1613	cyanide hydratase
7	523.5	28.7	346	2	S77025	nitriase (EC 3.5.
8	499.5	27.4	346	2	T49147	nitriase (EC 3.5.
9	493.5	27.0	346	2	S22398	nitriase (EC 3.5.
10	492.5	27.0	339	2	S31969	nitriase (EC 3.5.
11	491.5	26.9	339	2	T52259	nitriase (EC 3.5.
12	490.5	26.9	346	2	T49148	nitriase (EC 3.5.
13	489.5	26.8	339	2	T52262	nitriase (EC 3.5.
14	453.5	24.8	305	2	T27679	probable nitriase

15	451.5	24.7	362	2	T52266	nitriase-like pro
16	445	24.4	348	2	T03739	nitriase (EC 3.5.
17	443.5	24.3	355	2	T52265	nitriase (EC 3.5.
18	427.5	23.4	349	2	T03736	nitriase (EC 3.5.
19	257.5	14.1	199	2	S50363	nitriase homolog
20	207	11.3	272	2	T41662	probable nitriase
21	185	10.1	297	2	F75263	probable hydrolase
22	183	10.0	262	2	C71109	hypothetical prote
23	178.5	9.8	292	2	C87275	hypothetical prote
24	172.5	9.4	318	2	T48563	hypothetical prote
25	171	9.4	262	2	C75051	hydrolase related
26	160	8.8	257	2	C69264	conserved hypotet
27	156.5	8.6	579	2	AH3225	amidohydrolase lim
28	152.5	8.4	285	2	S58240	hypothetical prote
29	152.5	8.4	295	2	H82556	beta-alanine synth
30	151.5	8.3	322	2	T38399	probable amidohydr
31	150.5	8.2	576	2	G72277	probable amidohydr
32	146.5	8.0	267	2	B72408	conserved hypotet
33	143	7.8	291	2	S51459	hypothetical prote
34	143	7.8	294	2	AB0115	probable carbon-ni
35	142.5	7.8	292	2	E64558	conserved hypotet
36	142	7.8	312	2	TJW083	N-carbamyl-D-amino
37	140.5	7.7	294	2	G71949	hypothetical prote
38	138.5	7.6	272	2	B69109	N-carbamoyl-D-amin
39	138	7.6	280	2	T34905	probable hydrolase
40	136.5	7.5	393	2	S27881	beta-alanine synth
41	136	7.4	298	2	T17568	hydrolase homolog
42	135.5	7.4	290	2	B81369	probable hydrolase
43	134	7.3	220	2	A84673	probable nitriase
44	134	7.3	276	2	T40601	putative nitriase
45	132.5	7.3	292	2	G83608	probable hydratase

ALIGNMENTS

RESULT 1  
A45070 nitriase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A45070  
R/Kobayashi, M.; Komeda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
J. Biol. Chem. 267, 20746-20751, 1992  
A/Title: Nitriase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the  
A/Reference number: A45070; MUID:93015976; PMID:1400390  
A/Contents: J1  
A/Accession: A45070  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-366 <KOB>  
A/Cross-references: UNIPROT:Q03217; UNIPARC:UPI000016FF36; GB:D11425; NID:g216933; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:116039, NCBIP:116040)  
C/Superfamily: nitriase (carbon-nitrogen hydrolase)

Alignment Scores:  
Pred. No.: 6.11e-71  
Score: 836.50  
Percent Similarity: 68.8%  
Best Local Similarity: 50.6%  
Query Match: 45.8%  
DB: 2  
Length: 366  
Matches: 156  
Conservative: 56  
Mismatch: 95  
Indels: 1  
Gaps: 1

US-09-751-299-3 (1-1014) x A45070 (1-366)  
QY 16 AAGTCGCTGCGTCGACGCGCCCGCATCTACATGATTGGAGCGACGTGACAA 75  
Db 9 LysValAlaIaValGlnAlaGlnProValTrpPheAspAlaIaIaLysThrValAspLys 28  
QY 76 ACCATTGATTGATGAGAAGACGACGTAATATGCTGCTGATCGCCTTCCGAA 135  
Db 29 ThrValSerIleIleAlaGlnAlaIaIaArgAsnGlyCysGlnLeuValAlaPheProGlu 48  
QY 136 ACTTGATTCCAGGCTACCCATGTTCTTTGGCTTGACTCACGACATGGCAATG--- 192

Db 49 ValPheIleProGlyTyrProTyrHisIleTrrValAspSerProLeuAlaGlyMetAla 68  
:::|||||  
QY 193 CAATTGTAGCGCCAAATACCATGAGAACTCATGTGAGTTGGATGGCCCTCAAGCTTAAGCGC 252  
:::|||||  
Db 69 LysPheAlaValArgTyrHisGluAsnSerIleThrMetAspSerProHisValGlnArg 88  
:::|||||  
QY 253 ATTCAGATGCAAGCCAAAGCGGTTGGAAATCATGTGCACCCCTGGGATGAGTGAACGGGTC 312  
:::|||||  
Db 89 LeuLeuAspAlaAlaArgAspHisAsnIleAlaValValGlyIleSerGluArgAsp 108  
:::|||||  
QY 313 GGTGGCACCCCTTATCATCAGTCAGTGGTTCAATAGCCGATAATGTTGACACCATTTGGGGCC 372  
|||||  
Db 109 GlyGlySerLeuTyrMetThrGlnLeuValIleAspAlaAspGlyGlnLeuValAlaArg 128  
|||||  
QY 373 CGGCGAAAGTTGAAACCTACTTTGTGTAACGTACTTTGTTCCGGCGAAGGGGATGGTTCA 432  
|||||  
Db 129 ArgArgLysLeuLysProThrHisValGluArgSerValTyrGlyGluGlyAsnGlySer 148  
|||||  
QY 433 TCGCTAGCGGTTTTCGAGACGCTGTGTTGGAAGCGTGGGTGGCTTATGCTGTTGGGAGCAC 492  
:::|||||  
Db 149 AspIleSerValTyrAspMetProPheAlaArgLeuGlyAlaLeuAsnCysTrpGluHis 168  
|||||  
QY 493 CTTCACACCGCTTAACAATAATACGCTTTGTATGCACAAAATGAAGAGATTCATTTGCGGCT 552  
|||||  
Db 169 PheGlnThrLeuThrLysTyrAlaMetTyrSerMetHisGluGlnValHisValAlaSer 188  
|||||  
QY 553 TGGCCGAGCTTTAGCCTTTATCCTAATGCGGCAAGCCCTGGGCTGATGTCAATGTA 612  
|||||  
Db 189 TrpProGlyMetSerLeuTyrGlnProGluValProAlaPheGlyValAspAlaGlnLeu 208  
|||||  
QY 613 GCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTACTAGCGCTGTGCGCTC 672  
|||||  
Db 209 ThrAlaThrArgMetTyrAlaLeuGluGlyGlnThrPheValValCysThrThrGlnVal 228  
|||||  
QY 673 GTTTCACAATCCATGATCGATATGCTTGTGTACAGATGACGAAAGCATGCCTTGCTCTG 732  
|||||  
Db 229 ValThrProGluAlaHisGluPhePheCysAspAsnAspGluGlnArgLysLeuIleGly 248  
|||||  
QY 733 GCTGTGTGTGACACTCAGCTATCATAGGGCCCTGATGGTGTGACTTGGTCGCGCTCTT 792  
|||||  
Db 249 ArgGlyGlyGlyPheAlaArgIleIleGlyProAspGlyArgAspLeuAlaThrProLeu 268  
|||||  
QY 793 GCCGAAATGAAGGGTATTTCTCTACGCAAACTTGATCCTGTGAGTAGCGCATCTTGCT 852  
|||||  
Db 269 AlaGluAspGluGluGlyIleLeuTyrAlaAspIleAspLeuSerAlaIleThrLeuAla 288  
|||||  
QY 853 AAAATGGCGGAGACCCCTGCTGTCATTATTCGCCCTCCGACATTACTCGCTTGCTAATA 912  
|||||  
Db 289 LysGlnAlaAlaAspProValGlyHisTyrSerArgProAspValLeuSerLeuAsnPhe 308  
|||||  
QY 913 GATCGCAGCCCTAAATTACCGGTA 936  
:::|||||  
Db 309 AsnGlnArgHisThrThrProVal 316  
:::|||||  
RESULT 2  
A47181  
nitriIase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis  
C/Species: Alcaligenes faecalis  
C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A47181; S13860  
R/Kobayashi, M.; Izui, H.; Nagasawa, T.; Yamada, H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 247-251, 1993  
A/Title: NitriIase in biosynthesis of the plant hormone indole-3-acetic acid from indole  
A/Reference number: A47181; MUID:93126352; PMID:8419930  
A/Contents: JM3  
A/Accession: A47181  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-356 <KOB>  
A/Cross-references: UNIPROT:P20960; UNIPARC:UPI0000033514; GB:D13419; NID:g216202; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:122081, NCBI:P:122082)  
R/Nagasawa, T.; Manger, J.; Yamada, H.

Eur. J. Biochem. 194, 765-772, 1990  
A/Title: A novel nitriIase, arylacetone nitriIase, of Alcaligenes faecalis JM3. Purificatio  
A/Reference number: S13860; MUID:91099356; PMID:2269298  
A/Accession: S13860  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-33 <NAG>  
A/Cross-references: UNIPARC:UPI0000175E31  
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Alignment Scores:

Pred. No.:	2.77e-69	Length:	356
Score:	819.00	Matches:	156
Percent Similarity:	66.9%	Conservative:	54
Best Local Similarity:	49.7%	Mismatches:	104
Query Match:	44.9%	Indels:	0
DB:	2	Gaps:	0

US-09-751-299-3 (1-1014) x A47181 (1-356)

QY 4 AAAGAAGCTATCAAGGTGCGCTGCGTGCAGACCCCGCATCTACATGATTTGGAGCG 63  
:::|||||  
Db 4 ArgLysIleValArgAlaAlaAlaValGlnAlaAlaSerProAsnTyrAspLeuAlaThr 23  
|||||  
QY 64 ACGGTGACAAACCATTTGAGTTGATGGAAGAAGCAGACAGCATTAATGCTCGTGTGATC 123  
|||||  
Db 24 GlyValAspLysThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCysAspLeuIle 43  
|||||  
QY 124 GCCTTCCGGAACCTTGATTCAGGCTACAGCCATCGGTTTCTTGGCTTGACTCACCAGCA 183  
|||||  
Db 44 ValPheGlyGluThrTrpLeuProGlyTyrProPheHisValTrpLeuGlyAlaProAla 63  
|||||  
QY 184 TGGCAATGCAATTGTGACGCCAATACATGAGAACTCATGTGAGTTGATGGCCCTCAA 243  
|||||  
Db 64 TrpSerLeuLysTyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGlu 83  
|||||  
QY 244 GCTAAGCGCATTTCAATGCAGGCCCAAGCGGTTGGGAATCATGTGTCACCCCTGGGATGAGT 303  
|||||  
Db 84 PheGlnArgIleAlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSer 103  
|||||  
QY 304 GAACGGGCTCGGTGGCAACCTTTACATCATGATGCTTCATAGGCGCATTAATGTTGACACC 363  
|||||  
Db 104 GluArgSerGlyGlySerLeuTyrLeuGlyGlnCysLeuIleAspAspLysGlyGlnMet 123  
|||||  
QY 364 ATTGGGGCCCGGGAAGTTGAAACCTACTTTGTGTAACGTACTTTGTTCGGCGGAAGG 423  
:::|||||  
Db 124 LeuTrpSerArgArgLysLeuLysProThrHisValGluArgThrValPheGlyGlyGly 143  
|||||  
QY 424 GATGTTCAATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGCGTGGGCTTATGCTGT 483  
|||||  
Db 144 TyrAlaArgAspLeuIleValSerAspThrGluLeuGlyArgValGlyAlaLeuCysCys 163  
|||||  
QY 484 TGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAGAGATTTCAT 543  
|||||  
Db 164 TrpGluHisLeuSerProLeuSerLysTyrAlaLeuTyrSerGlnHisGluAlaIleHis 183  
|||||  
QY 604 GTCAATGTAGCGGCTCTCGAATCTATGCCCTTGAAGGCAATGCTTCGTACTAGCGTGC 663  
|||||  
Db 204 ValAsnMetAlaAlaSerGlnIleTyrSerValGluGlyGlnCysPheThrIleAlaAla 223  
|||||  
QY 664 TGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCG 723  
:::|||||  
Db 224 SerSerValValThrGlnGluThrLeuAspMetLeuGluValGlyGluHisAsnAlaSer 243  
|||||  
QY 724 TTGCTTGTGCGTGTGTGGACACTCAGTAATCATAGGGCGCTGATGGTGTGACTTGTGTC 783  
|||||  
Db 244 LeuLeuLysValGlyGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAla 263  
|||||

```
OY 784 GCGCCTTTCGCCGAAATGAGAGGGTATCTCTACGCAACCTTGATCCTGGAGTACGC 843
    |||
    ::|
Db 264 ProTyrLeuProHisAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluGluIle 283
OY 844 ATCCCTGTAATAATGCGCGCAGACCCCTGCTGTGTCATTATCCCGTCCGACATTACTCGC 903
    |||
    |||
Db 284 AlaPheAlaLysAlaIleAsnAspProValGlyHisTyrSerLysProGluAlaThrArg 303
OY 904 TTGCTAATAGATCGCAGCCCTAAATACCGGTAGTTGAATT 945
    |||
    |||
Db 304 LeuValLeuAspLeuGlyHisArgGluProMetThrArgVal 317
```

RESULT 3  
A43470  
aliphatic nitrilase - Rhodococcus rhodochrous  
C:Species: Rhodococcus rhodochrous  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C:Accession: A43470  
R:Kobayashi, M.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
Biochemistry 31, 9000-9007, 1992  
A:Title: Primary structure of an aliphatic nitrile-degrading enzyme, aliphatic nitrilase  
A:Reference number: A43470; MUID:93003039; PMID:1390687  
A:Contents: K22  
A:Accession: A43470  
A>Status: preliminary  
A:Molecule type: DNA; protein  
A:Residues: 1-383 <KOB>  
A:Cross-references: UNIPROT:Q02068; UNIPARC:UPI0000130516; GB:D12583; NID:g216931; PIDN:  
A:Note: sequence extracted from NCBI backbone (NCBIN:114184, NCBIP:114185)  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

Alignment Scores:  
Pred. No.: 1.79e-67 Length: 383  
Score: 800.00 Matches: 159  
Percent Similarity: 66.5% Conservative: 53  
Best Local Similarity: 49.8% Mismatches: 103  
Query Match: 43.8% Indels: 4  
DB: 2 Gaps: 4

US-09-751-299-3 (1-1014) x A43470 (1-383)

```
OY 13 ATCAAGTCGCGCTGCGTCAAGCGCCCGATCTACATGATTTGAGCGCAGCGGTGAC 72
    ::|
    |||
Db 13 ValLysValAlaThrValGlnAlaGluProValIleLeuAspAlaAspAlaThrIleAsp 32
OY 73 AAACCATGTAGTTGATGAAGAAGCAGCAGCATATAATGCTCGTCTGATCGCCTTCCG 132
    |||
    ::|
Db 33 LysAlaIleGlyPheIleGluGluAlaAlaLysAsnGlyAlaGluPheLeuAlaPhePro 52
OY 133 GAAACTTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCACCAAGCATGGGCAATG 192
    |||
    |||
Db 53 GluValTrpIleProGlyTyrProTyrTrpAlaTrpIleGlyAspValLysTrpAlaVal 72
OY 193 ---CAATTGTACGCCAATACCATGAGAATCTCATTTGAGTTGATGGCCCTCAAGCTAAG 249
    |||
    |||
Db 73 SerAspPheIleProLysTyrHisGluAsnSerLeuThrLeuGlyAspAspArgMetArg 92
OY 250 CGCATTTACAGATGCAGCCCAAGCGGTGGGAATCATGTGTCACCTGGGATGAGTGAACG 309
    |||
    |||
Db 93 ArgLeuGlnLeuAlaAlaArgGlnAsnAsnIleAlaLeuValMetGlyTyrSerGluLys 112
OY 310 GTCGGTGGCACCCTTTACATCAGTCAGTGTTCATAGCGGATAATGTTGACACCAATGGG 369
    |||
    ::|
Db 113 AspGlyAlaSerArgTyrLeuSerGlnValPheIleAspGlnAsnGlyAspIleValAla 132
OY 370 GCCCGCGGAAAGTTGAACCTACTTTTGTGACAGTACTTTGTCGGCGAAGGGATGCT 429
    |||
    |||
Db 133 AsnArgArgLysLeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGly 152
OY 430 TCATCGCTAGCGGTTTTCGAGACGCTCTGTTGGAAGCGTCGGCTTATGCTCTTGGAG 489
    ::|
    ::|
Db 153 ThrAspPheLeuThrHisAspPheGlyPheGlyArgValGlyGlyLeuAsnCysTrpGlu 172
```

```
OY 490 CACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGATTCATTGTGCG 549
    |||
    |||
Db 173 HisPheGlnProLeuSerLysTyrMetMetTyrSerLeuAsnGluGlnIleHisValAla 192
OY 550 GCTTGGCCGAGC--TTTAGCCTTATCCTAATAGCGGCGAAAGCCCTGGGCGCTGATGTC 606
    ::|
    |||
Db 193 SerTrpProAlaMetPheAlaLeuThrProAspValHisGln---LeuSerValGluAla 211
OY 607 AATGACCGGCTCTCGAATCTATGCCGTTGAAGGCGCAATGCTTCTACTAGCGTCTGT 666
    |||
    |||
Db 212 AsnAspThrValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuAlaSerThr 231
OY 667 GCGCTGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAA--AAGCATGCG 723
    ::|
    |||
Db 232 HisValIleGlyLysAlaThrGlnAspLeuPheAlaGlyAspAspAlaLysArgAla 251
OY 724 TTGCTTCTGGCTGTGTGGACACTCACGTCATCATAGGCGCTGATGCTGTGACTTGTGTC 783
    |||
    |||
Db 252 LeuLeuProLeuGlyGlnGlyTrpAlaArgIleTyrGlyProAspGlyLysSerLeuAla 271
OY 784 GCGCCTTTCGCCGAAATGAGAGGGTATCTCTACGCAACCTTGATCCTGAGTACGC 843
    |||
    |||
Db 272 GluProLeuProGluAspAlaGluGlyLeuLeuTyrAlaGluLeuAspLeuGluGlnIle 291
OY 844 ATCCCTGTAATAATGCGCGCAGACCCCTGCTGTGTCATTATCCCGTCCGACATTACTCGC 903
    |||
    |||
Db 292 IleLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgProAspValLeuSer 311
OY 904 TTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATTGAAGGTGATCTTCGT 960
    |||
    |||
Db 312 LeuLysIleAspThrArgAsnHisThrProValGlnTyrIleThrAlaAspGlyArg 330
```

RESULT 4  
JC4212  
nitrilase (EC 3.5.5.1) - Comamonas testosteroni  
C:Species: Comamonas testosteroni  
C:Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 31-Dec-2004  
C:Accession: JC4212; PC4056  
R:Levy-Schil, S.; Soubrier, F.; Crutlz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D.  
Gene 161, 15-20, 1995  
A:Title: Aliphatic nitrilase from a soil-isolated Comamonas testosteroni sp.: Gene clone  
A:Reference number: JC4212; MUID:95369726; PMID:7642130  
A:Accession: JC4212  
A:Molecule type: DNA  
A:Residues: 1-354 <LEV>  
A:Cross-references: UNIPROT:Q59329; UNIPARC:UPI00000B98BB; GB:L32589; NID:g1082008; PIDN:  
A:Accession: PC4056  
A:Molecule type: protein  
A:Residues: 154-162;295-302;323-340 <LE2>  
A:Cross-references: UNIPARC:UPI0000175B32; UNIPARC:UPI0000175E33; UNIPARC:UPI0000175E34  
A:Comment: This enzyme is active on adiponitrile and cyanovaleric acid.  
C:Genetics:  
A:Gene: nita  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase  
F:163/Active site: Cys #status predicted

Alignment Scores:  
Pred. No.: 4.07e-52 Length: 354  
Score: 638.00 Matches: 134  
Percent Similarity: 60.3% Conservative: 62  
Best Local Similarity: 41.2% Mismatches: 107  
Query Match: 34.9% Indels: 22  
DB: 2 Gaps: 5

US-09-751-299-3 (1-1014) x JC4212 (1-354)

```
OY 13 ATCAAGTCGCGCTGCGTCAAGCGCCCGATCTACATGATTTGAGCGCAGCGGTGAC 72
    ::|
    |||
Db 7 ValLysValAlaAlaValGlnAlaAlaProValPheMetAsnLeuGluAlaThrValAsp 26
OY 73 AAACCATGTAGTTGATGAAGAAGCAGCAGCATTAATGCTCTGCTGATCGCCTTCCG 132
    |||
    ::|
```



Db	27	LysThrCysLysLeuIleAlaGluAlaLaserMetGlyAlaLysValIleGlyPhePro	46
QY	133	GAAACTTGATTCAGGCTAACCCATGGTTCTTGCGCTGACTCA-----	177
Db	47	GluAlaPheIleProGlyTyrProTyrTrpIleTrpThrSerAsnMetAspPheThrGly	66
QY	178	CCAGCATGGGCAATGCATTGTGTACGCCAATACCATGAGAATCATGTGAGTTGGATGGC	237
Db	67	MetMetTrpAlaValLeuPhe-----LysAsnAlaIleGluIleProSer	81
QY	238	CCTCAAGCTAAGCCGATTTCAAGATGCAGCCCAAGCGGTTGGGAATCATGTGCACCCCTGGG	297
Db	82	LysGluValGlnGlnIleSerAspAlaAlaLysLysAsnGlyValTyrValCysValSer	101
QY	298	ATGAGTGAAACGGGTCCGTGGTGACCCCTTACATCAGTCAGTGGTTCATAGGCGATAATGGT	357
Db	102	ValSerGluLysAspAsnAlaSerLeuTyrLeuThrGlnLeuTrpPheAspProAsnGly	121
QY	358	GACACCATTTGGGCGCCCGGAAAGTTGAAACCTACTTTGTGAAACGTACTTTGTTGGC	417
Db	122	AsnLeuIleGlyLysHisArgLysPheLysProThrSerSerGluArgAlaValTrpGly	141
QY	418	GAAAGGAGTGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTTGGAAGCGTGGTGCTTA	477
Db	142	AspGlyAspGlySerMetAlaProValPheLysThrGluTyrGlyAsnLeuGlyGlyLeu	161
QY	478	TGCTGTGGAGACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAG	537
Db	162	GlnCysTrpGluHisAlaLeuProLeuAsnIleAlaIleMetGlySerLeuAsnGluGln	181
QY	538	ATTCAATTGTGCGGCTTGCGCGAGCTTTAGCCCTTATCCTAATGCGCGAAAGCCCTGGG	597
Db	182	ValHisValAlaSerTrpProAlaPhe-----ValProlysGlyAlaValSerSerArg	199
QY	598	CCTGATGTCAATGTAGCGGC-----TCTGGAATCTAT	630
Db	200	ValSerSerSerValCysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyr	219
QY	631	GCCGTTGAAGGGCAATGCTTCGTACTAGCGTCGTGCGCTCGTTTCACAATCCATGATC	690
Db	220	AlaIleSerAsnGlnValTyrValIleMetSerThrAsnLeuValGlyGlnAspMetIle	239
QY	691	GATATGCTTTGTACAGATGACGAAAGCATGCGTTCCTTGCTGGCTGGTGATGACACTCA	750
Db	240	AspMetIleGlyLysAspGluPheSerLysAsnPheLeuProLeuGlySerGlyAsnThr	259
QY	751	CGTATCATAGGGCTGATGGTGGTGAATGTCGCGCTCTTGCCGAAATGAAGAGGT	810
Db	260	AlaIleIleSer--AsnThrGlyGluIleLeuAlaSerIleProGlnAspAlaGluGly	278
QY	811	ATTCTCTACGCAACCTTGATCCTGAGATACGCATCCTTGCTAAATGGCGCAGACCCCT	870
Db	279	IleAlaValAlaGluIleAspLeuAsnGlnIleIleTyrGlyLysTrpLeuAspPro	298
QY	871	GCTGTCATTATTCCTCCGTCGCGACATTACTCGCTTGCTAATAGATCGACGCCCTAAATT	930
Db	299	AlaGlyHisTyrSerThrProGlyPheLeuSerLeuThrPheAspGlnSerGluHisVal	318
QY	931	CCGGTAGTTGAATT	945
Db	319	ProValLysLysIle	323
RESULT 5			
A28658			
nitriIase (EC 3.5.5.1) - Klebsiella ozaenae			
C/Species: Klebsiella ozaenae			
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Dec-2004			
C/Accession: A28658			
R/Stalker, D.M.; Malyj, L.D.; McBride, K.E.			
J. Biol. Chem. 263, 6310-6314, 1988			
A/Title: Purification and properties of a nitriIase specific for the herbicide bromoxyni			
A/Reference number: A28658; MUID:88198177; PMID:2834373			
< A/Accession: A28658			

A/Molecule type: DNA			
A/Residues: 1-349 <STA>			
A/Cross-references: UNIPROT:P10045; UNIPARC:UPI000013051A; GB:J03196; NID:g149174; PIDN:f			
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)			
C/Keywords: hydrolase			
Alignment Scores:			
Pred. No.:			
Score:			
Percent Similarity:			
Best Local Similarity:			
Query Match:			
DB:			
US-09-751-299-3 (1-1014) x A28658 (1-349)			
QY	1	ATGAAGAAGCTATCAAGGTCGCGCTGCGAAGCCCGCCGATCTACATGATTGGAG	60
Db	1	MetAspThrThrPheLysAlaAlaAlaValGlnAlaGluProValTrpMetAspAlaAla	20
QY	61	GCGACGGTGGACAAACCATTTGATGATGGAAGAAGCAGCAGCATTAATGATGCTGCTG	120
Db	21	AlaThrAlaAspLysThrValThrLeuValAlaLysAlaAlaAlaGlyAlaGlnLeu	40
QY	121	ATCGCCTTCCGGAACCTTGATTCAGGCTAACCCATGTTCTTGGCTGACTCACCA	180
Db	41	ValAlaPheProGluLeuTrpIleProGlyTyrProGlyPheMetLeuThrHisAsnGln	60
QY	181	GCATGGGCAATGCAATTTGTACGCCAATGCCATGAGAATCATGTGAGTGGAGCCCT	240
Db	61	ThrGluThrLeuProPheIleIleLysTyrArgLysGlnAlaIleAlaAspGlyPro	80
QY	241	CAAGCTAAGCGCATTTCAAGTGCAGCGCAAGCGGTTGGAAATCATGGTCACCCTGGGATG	300
Db	81	GluIleGluLysIleArgCysAlaAlaGlnGluHisAsnIleAlaLeuSerPheGlyTyr	100
QY	301	AGTGAACGGGTGCGTGGACACCCCTTACATCAGTCAGTGGTTCATAGGCGAATAGTGAC	360
Db	101	SerGluArgAlaGlyArgThrLeuTyrMetSerGlnMetLeuIleAspAlaAspGlyIle	120
QY	361	ACCATTGGGGCGCCGCGAAAGTTGAAACCTATTGTTGAACGTACTTTGTCGGCGAA	420
Db	121	ThrLysIleArgArgArgLysLeuLysProThrArgPheGluArgGluLeuPheGlyGlu	140
QY	421	GCGGATGTTTCATCGCTAGCGGTTTTCGAGAGCTGTGGAAGCGTGGGCTTATGC	480
Db	141	GlyAspGlySerAspLeuGlnValAlaGlnThrSerValGlyArgValGlyAlaLeuAsn	160
QY	481	TGTTGGAGACCTTCAACCGCTAACAAATAGCCTTGTATGCACAAATGAAGAGATT	540
Db	161	CysAlaGluAsnLeuGlnSerLeuAsnLysPheAlaLeuAlaGluGlyGlnGlnIle	180
QY	541	CATTGCGGCTTGCGCGAGCTTTAGCCTTATCCTAATGCGCGAAAGCCCTGGGGCT	600
Db	181	HisIleSerAlaTrpPro--PheThrLeu-----GlySerProValLeuValGly	196
QY	601	GATGTCAATGAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTGCTAGCGG	660
Db	197	AspSerIleGlyAlaIleAsnGlnValTyrAlaAlaGluThrGlyThrPheValLeuMet	216
QY	661	TGCTGTGCGCTGTTTCACAATCCATGATGCATATGCTTTGTACAGATGACGAAAGCAT	720
Db	217	SerThrGlnValGlyProThrGlyIleAlaAlaPheGluIleGluAspArgTyrAsn	236
QY	721	GCGTGTCTTGCGCTGGTGGTGAGACACTCAGTATCATAGGCGCTGATGGTGACTTG	780
Db	237	ProAsnGlnTyrLeuGlyGlyGlyTyrAlaArgIleTyrGlyProAspMetClnLeuLys	256
QY	781	GTCGCGCTTTCGCCGAAATGAAGAGGATTTCTTACGCAACCTTGATCTGAGTA	840
Db	257	SerLysSerLeuSerProThrGluGluGlyIleValTyrAlaGluIleAspLeuSerMet	276
QY	841	CGCATCCTTGCTAAATGGCGGCGAGACCCCTGCTGTCATTATTCGCCGTCGACATTACT	900

Db 277 LeuGluAlaAlaLysTyrSerLeuAspProThrGlyHisTyrSerArgProAspValPhe 296  
QY 901 CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAAATT-----GAAGT 951  
Db 297 SerValSerIleAsnArgGlnArgGlnProAlaValSerGluValIleAspSerAsnGly 316  
QY 952 GATCTTCGTCCTTACGCTTTGGGTAAAGCGTCTGAG 987  
Db 317 AspGluAspProArgAlaAlaCysGluProAspGlu 328  
RESULT 6  
JQ1613  
cyanide hydratase (EC 4.2.1.66) - imperfect fungus (Gloeocercospora sorghi)  
C/Species: Gloeocercospora sorghi  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004  
C/Accession: JQ1613  
R/Wang, P.; VanEtten, H.D.  
Biochem. Biophys. Res. Commun. 187, 1048-1054, 1992  
A/Title: Cloning and properties of a cyanide hydratase gene from the phytopathogenic fun  
A/Reference number: JQ1613; MUID:92412068; PMID:1382413  
A/Accession: JQ1613  
A/Molecule type: DNA  
A/Residues: 1-368 <WAN>  
A/Cross-references: UNIPROT:P32964; UNIPARC:UPI0000128C2E; GB:M99044; NID:gl68196; PIDN:  
C/Comment: This enzyme converts HCN to formamide.  
C/Genetics:  
A/Gene: Cht  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase  
F/286/Binding site: carbohydrate (asn) (covalent) #status predicted  
Alignment Scores:  
Pred. No.: 1.28e-44 Length: 368  
Score: 559.00 Matches: 128  
Percent Similarity: 56.1% Conservative: 56  
Best Local Similarity: 39.0% Mismatches: 118  
Query Match: 30.6% Indels: 26  
DB: 2 Gaps: 7  
US-09-751-299-3 (1-1014) x JQ1613 (1-368)  
QY 16 AAGTCGCTGCGTGCAGCCGCCCGATCTACATGATTGGAGCGGCGGTGACAAA 75  
Db 7 LysAlaAlaValValThrSerGluProValTrpGluAsnGluGlyGlyValVallys 26  
QY 76 ACCATTGAGTTGATGGAAGACGACGACGTAATAAGTCGTCGTGATCGCCTTCCGAA 135  
Db 27 ThrIleGluPheIleAsnGluAlaGlyLysAlaGlyCysLysLeuIleAlaPheProGlu 46  
QY 136 ACTTGATTCCAGGCTACCGCATGTTCTTTGGCTTGACTCACACGATGGCAATGCAA 195  
Db 47 ValTrpIleProGlyTyrProTyrTrpMetTrpLysValAsnTyrLeuGlnSerLeuPro 66  
QY 196 TTTGTAGCCCATATACCATGAGAATCATGTGAGTTGATGGCCCTCAAGCTAAGCGCAT 255  
Db 67 MetLeuLysAlaTyrArgGluAsnSerIleAlaMetAspSerSerGluMetArgArgIle 86  
QY 256 TCAGATGCAGCCAGCCGTTGGGAATCATGTGTCACCCCTGGGGATGATGAACGGGTCGGT 315  
Db 87 ArgAlaAlaAlaArgAspAsnGlnIleTyrValSerIleGlyValSerGluIleAspHis 106  
QY 316 GGCACCCCTTACATCAGTACGTGTTTCATAGGGCGATAATGTTGACACCATTTGGGCGCG 375  
Db 107 AlaThrLeuTyrLeuThrGlnValLeuIleSerProLeuGlyAspValIleAsnHisArg 126  
QY 376 CGAAAGTTGAAACCTACTTTTGTGAACGTAATTGTTGGCGAAGGGAGTTCATCG 435  
Db 127 ArgLysIleLysProThrHisValGluLysLeuValTyrGlyAspGlySerGlyAspSer 146  
QY 436 CTA---GCGGTTTTCAGACGTCGTGTTGGAAGGCTGGGCTTATGCTGTTGGAGCAC 492  
Db 147 PheGluProValThrGlnThrGluIleGlyArgLeuGlyGlnLeuAsnCysTrpGluAsn 166

QY 493 CTTCAACCGCTAACAATAACGCTTTGTATGCACAATAATGAAGATTCATTGTGCGCT 552  
Db 167 MetAsnProPheLeuLysSerLeuAlaValAlaArgGlyGluGlnIleHisValAlaIa 186  
QY 553 TGGCCGAGCTTTAGCCTTTATCTTAATGCGCGGAAGCCCTGGGCGCTGAT-----GTC 606  
Db 187 TrpPro-----ValTyrProAspLeuSerLysGlnValHisProAspProAlaThr 203  
QY 607 AATGTAGCG-----GCTCTCGAATCTATGCCGTTGAAGGCAATGC 648  
Db 204 AsnTyrAlaAspProAlaSerAspLeuValThrProAlaTyrAlaIleGluThrGlyThr 223  
QY 649 TTCGTACTAGCGTCGTGTCGCTCGTTTCACAATCCATGATCGATATGCTTGT----- 702  
Db 224 TrpValLeuAlaProPheGlnArgIleSerValGluGlyLeuLysArgHisThrProPro 243  
QY 703 -----ACAGATGACGAAAGCATGCCGTTCTTGCGCTGGTGGACAC 747  
Db 244 GlyValGluProGluThrAspAlaThrProTyr-----AsnGlyHis 257  
QY 748 TCACGTATCATAGGCGCTGATGTTGTTGCTGCTGCGCGCTTGGCCGAAATGAAGAG 807  
Db 258 AlaArgIlePheArgProAsp--GlySerLeuTyrAlaLysProAlaValAspPheAsp 276  
QY 808 GGTATTCTCTACGAAACCTTGATCCTGAGTACGCATCCTTGCTAAATGCGCGAC 867  
Db 277 GlyLeuMetTyrValAspIleAspLeuAsnGluSerHisLeuThrLysAlaLeuAlaAsp 296  
QY 868 CCTGCTGTATATTCCCGTCCCGACATGATCTGCTGCTAATAGATCGAGCCCTAAA 927  
Db 297 PheAlaGlyHisTyrMetArgProAspLeuIleArgLeuLeuValAspThrArgArgLys 316  
QY 928 TTACCGGTAGTTGAATTGAAGT 951  
Db 317 GluLeuValThrGluValGlyGly 324  
RESULT 7  
S77025  
nitrilase (EC 3.5.5.1) - Synechocystis sp. (strain PCC 6803)  
N/Alternate names: protein gl10784  
C/Species: Synechocystis sp.  
A/Variety: PCC 6803  
C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004  
C/Accession: S77025  
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,  
DNA Res. 3, 109-136, 1996  
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
S.  
A/Reference number: S74322; MUID:97061201; PMID:8905231  
A/Accession: S77025  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-346 <KAN>  
A/Cross-references: UNIPROT:Q55949; UNIPARC:UPI00000D7133; EMBL:D64005; GB:AB001339; NID:  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C/Genetics:  
A/Start codon: GTG  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Alignment Scores:  
Pred. No.: 2.93e-41 Length: 346  
Score: 523.50 Matches: 109  
Percent Similarity: 58.0% Conservative: 68  
Best Local Similarity: 35.7% Mismatches: 119  
Query Match: 28.7% Indels: 9  
DB: 2 Gaps: 4  
US-09-751-299-3 (1-1014) x S77025 (1-346)  
QY 13 ATCAAGTGCCTGCGTGCAGCGCCCGCATCTACATGATTTGGAGCGACGCTGAC 72

Db 13 ILeArgAlaAlaAlaGlnIleSerProValLeuPheSerGlnGlnGlyThrMetGlu 32  
QY 73 AAACCATTTGAGTTGATGGAAGAAGACGACGACGTAATATGCTGCTGTGATCGCCTTTCCG 132  
Db 33 LysValLeuAspAlaIleAlaAsnAlaAlaIleLysGlyValGluLeuIleValPhePro 52  
QY 133 GAAACTTGATTCAGGCTACCCATGCTGTCTTGCTTGACTCACCAGCATGGGCAATG 192  
Db 53 GluThrPheValProIleTyrProTyrPheSerPheValGluProProValLeuMetGly 72  
QY 193 CAATTTGTACGCCAATACCATGAGAAGACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCCG 252  
Db 73 LysSerHisLeuLysLeuTyrGlnGluAlaValThrValProGlyLysValThrGlnAla 92  
QY 253 ATTCAGATGCAGCCACGCGGTTGGGAATCATGTGTCACCCTGGGATGAGTGAACGGGTC 312  
Db 93 IleAlaGlnAlaAlaLysThrHisGlyMetValValLeuGlyValAsnGluArgGlu 112  
QY 313 GGTGGCACCCTTTACATAGTCAGTGGTTCATAGGCGATATATGTGTGACACCATTTGGGGCC 372  
Db 113 GluGlySerLeuTyrAsnThrGlnLeuIlePheAspAlaAspGlyAlaLeuValLeuLys 132  
QY 373 CGGCGAAAGTTGAACCTACTTTGTGTGAACGTAATTTGTCGGCGAAGGGAGTGGTTCA 432  
Db 133 ArgArgLysIleThrProThrTyrHisGlyArgMetValTyrGlyGlnGlyAspGlyAla 152  
QY 433 TCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGTGGCTTATGCTGTGGAGACAC 492  
Db 153 GlyLeuArgThrValAspThrThrValGlyArgLeuGlyAlaLeuAlaCysTyrGlnHis 172  
QY 493 CTTCACCGCTAACAAATACGCTTGTATGACACAAATGAAGAGATTCAATTGTGGGCT 552  
Db 173 TyrAsnProLeuAlaArgTyrAlaLeuMetAlaGlnHisGlyGlnIleHisCysGlyGln 192  
QY 553 TGGCCGAGCTTTAGCCTTTATCTTAATGCGGCGAAGCCCTGGGCTGATGTCAATGTA 612  
Db 193 PheProGly-----SerMetValGlyGlnIlePheAlaAspGlnMetGlu 207  
QY 613 GCGGCTCTCGAATCTATGCGCGTTGAAGGCGCAATGCTTCGTACTA--GCGTCGTGTGCG 669  
Db 208 ValThrMetArgHisHisAlaLeuGluSerGlyCysPheValIleAsnAlaThrGlyTyr 227  
QY 670 CTGCTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTTGCTT 729  
Db 228 LeuThrAlaGluGlnLysLeuGlnIle-----ThrThrAspGluLysMetHisGlnAla 245  
QY 730 CTGCGCTGTGTGTGACACTCACTATCATAGGCGCTGATGTGTGACTTGTGCGGCT 789  
Db 246 LeuSerGlyGlyCysTyrThrAlaIleIleSerProGluGlyLysHisLeuCysGluPro 265  
QY 790 CTTCGCCGAAAATGAGAGGGTATCTCTACGCAACCTTGATCCTGAGTACGCATCCTT 849  
Db 266 IleAlaGluGly--GluGlyLeuAlaIleAlaAspLeuAspPheSerLeuIleAlaLys 284  
QY 850 GCTTAAATGGCGGACAGACCCCTGCTGTCATTAATCCCGTCCGACATTAATCGCTTGCTA 909  
Db 285 ArgLysArgMetMetAspSerValGlyHisTyrAlaArgProAspLeuLeuGlnLeuThr 304  
QY 910 ATAGATCGCAGCCCT 924  
Db 305 LeuAsnAsnGlnPro 309

RESULT 8  
T49147  
nitrlase (EC 3.5.5.1) 1 [imported] - Arabidopsis thaliana  
N/Alternate names: protein T10D17.100  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 31-Dec-2004  
C/Accession: T49147; T52260; T52263  
R/D/Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L  
A/Reference number: Z25017

```

A/Accession: T49147
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-346 <DAN>
A/Cross-references: UNIPROT:P32961; UNIPARC:UPI0000001AE2; EMBL:AL353865; GSPDB:GN00061;
A/Experimental source: cultivar Columbia; BAC clone T10D17
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.
Plant Mol. Biol. 36, 89-99, 1998
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz
A/Reference number: Z26007; MUID:98145459; PMID:9484465
A/Accession: T52260
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-346 <HIL>
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAA68935.2
A/Experimental source: cultivar Columbia
R/Bartel, B.; Fink, G.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994
A/Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop
A/Reference number: Z24515; MUID:8022831; PMID:8022831
A/Accession: T52263
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-346 <BAR>
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:U38845; PIDN:AAB05221.1
A/Experimental source: cultivar Columbia
C/Genetics:
A/Gene: ATSP:T10D17.100; nit1; NIT1
A/Map position: 3
A/Intons: 44/1; 104/1; 202/1; 296/1
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase

Alignment Scores:
Pred. No.: 5.53e-39 Length: 346
Score: 499.50 Matches: 113
Percent Similarity: 54.3% Conservative: 65
Best Local Similarity: 34.5% Mismatches: 109
Query Match: 27.4% Indels: 41
DB: 2 Gaps: 8

US-09-751-299-3 (1-1014) x T49147 (1-346)

QY 13 ATCAAGTCGCTGCTGCGCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGAC 72
:::||||| |||||::: ||||| |||||::: |||||
25 ValArgValThrIleValGlnSerSerThrValTyrAsnAspThrProIaThrIleasp 44

QY 73 AAACCATTTGAGTTGATGGAAGAAGCAGCAGCAGTAATATGCTCGTGCCTTCCG 132
||| ::| ||||| ||||| ||||| |||||
45 LysAlaGluLysTyrIleValGluAlaIaSerLysGlyAlaGluLeuValLeuPhePro 64

QY 133 GAAACTTGATTCAGGCTACCCA-----TGTTTCTTTGGCTTGAC 174
||| ::||| ||||| ||||| ||||| |||||
65 GluGlyPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuAlaValGlyValHis 84

QY 175 TCACCAGCATGGGCATGCAATTTGTACGCCCATATACCATGAGACTCATTTGAGTTGGAT 234
::: ||| ::||| ||||| ||||| ||||| |||||
85 AsnGluGluGlyArgAspGluPhe---ArgLysTyrHisAlaSerAlaIleHisValPro 103

QY 235 GGGCCTCAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGGAATCATGTCACCCCTG 294
|||||::: |||||::: |||||::: |||||::: |||||::: |||||::: |||||
104 GlyProGluValAlaArgLeuAlaAspValAlaArgLysAsnHisValTyrLeuValMet 123

QY 295 GGGATGAGTGAACGGGCGGTGGTGGCACCCTTTACATCAGTCAGTGGTTTCATAGCGGATTAAT 354
||| |||::: ||| ||||| ||||| ||||| |||||
124 GlyAlaIleGlyLysGlyGlyTyrThrLeuTyrCysThrValLeuPhePheSerProGln 143

QY 355 GGTGAACACCATTTGGGCGCGCGGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTC 414
||| ||| ::||| ||||| ||||| ||||| ||||| |||||
144 GlyGlnPheLeuGlyLysHisArgLysLeuMetProThrSerLeuGluArgCysIleTrp 163

QY 415 GCGCAAGGGGATGTTTCACTGCGCTAGCGGTTTTCAGACGTCGTGTTGAAGCTGGGTGGC 474
|||::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```
Db      164 GLYGLNGLYAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGlyAla 183
QY      475 TTATGCTGTGGAGCAGCACCCTTCAACCGCTTAACAATAATACCTTTGTATGCACAATAATGA 534
Db      184 AlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaLysGlyIle 203
QY      535 GAGATTCATGTGCG-----GCTGGCCGAGCTTTAGCCTTAT 573
Db      204 GluLeuTyrCysAlaProThrAlaAspGlySerLysGluTrpGlnSerSerMetLeuHis 223
QY      574 CCTAATCGCGGAAGCCCTGGGCGCTGATGTCATGTAGCGGCGCTCTGAAATCTATGCC 633
Db      224 -----IleAla 225
QY      634 GTTGAAGGCAATGCTTCGTAAGCGTCGTGCGCTGTTTCACAATCCATGATC--- 690
Db      226 IleGluGlyGlyCysPheValLeuSerAlaCysGlnPheCysGlnArgLysHisPhePro 245
QY      691 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCGTGTGCTTCTG 732
Db      246 AspHisProAspTyrLeuPheThrAspTrpTyrAspAspLysGluHisAspSerIleVal 265
QY      733 GCTGGTGTGGACACTCACGTAATAGAGGCGCTGATGGTGTGACTTGTCGCGCCTTT 792
Db      266 SerGlnGlyGly--SerValIleIleSerProLeuGlyGlnValLeuAlaGlyProAsn 284
QY      793 GCCGAAATGAAGAGGTATTTCTTACGCAACCTTGATCCTGGAGTAGCATCCTTGCT 852
Db      285 PheGluSer--GluGlyLeuValThrAlaAspIleAspLeuGlyAspIleAlaArgAla 303
QY      853 AAAATGGCGGACAGCCCTGCTGTCATTATTCGCCGTCGACATTACTGCTGTGTAATA 912
Db      304 LysLeuTyrPheAspSerValGlyHisTyrSerArgProAspValLeuHisLeuThrVal 323
QY      913 GATCGCAGCCCTAAATTACCGGTA 936
Db      324 AsnGluHisProArgLysSerVal 331
```

```
RESULT 9
S22398
nitriIase (EC 3.5.5.1) - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C/Accession: S22398
R/Bartling, D.; Seedorf, M.; Mithoefer, A.; Weiler, E.W.
Eur. J. Biochem. 205, 417-424, 1992
A/Title: Cloning and expression of an Arabidopsis nitriIase which can convert indole-3-a
A/Reference number: S22398; MUID:92209532; PMID:1555601
A/Accession: S22398
A/Molecule type: mRNA
A/Residues: 1-346 <BAR>
A/Cross-references: UNIPROT:P32961; UNIPARC:UPI0000130513; EMBL:X63445; NID:g16399; PIDN
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase
```

```
Alignment Scores:
Pred. No.:      2.05e-38      Length:      346
Score:          493.50       Matches:      112
Percent Similarity: 54.3%     Conservative: 66
Best Local Similarity: 34.1%   Mismatches:   109
Query Match:    27.0%        Indels:        41
DB:             2           Gaps:          8
```

```
US-09-751-299-3 (1-1014) x S22398 (1-346)
QY      13 ATCAAGTCGCGCTGCGTCAAGCCCGCCGATCTACATGATTTGGAGGCGAGCGGTGAC 72
Db      25 ValArgValThrIleValGlnSerSerThrValTyrAsnAspThrProAlaThrIleAsp 44
QY      73 AAAACCATGTAGTTGATGGAAGACGACGACGTAATAATGCTCGTCTGATCGCCTTTCCG 132
Db      45 LysAlaGluLysTyrIleValGluAlaAlaSerLysGlyAlaGluLeuValLeuPhePro 64
```

```
QY      133 GAAACTTGATTCAGGCTAACCA-----TGCTTCTTTGGCTTGAC 174
Db      65 GluGlyPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuAlaValGlyValHis 84
QY      175 TCACCAGCATGGGCAATGCAATTGTGTACGCCAATACCAATGAACTCATTTGAGTTGGAT 234
Db      85 AsnGluGlyGlyArgAspGluPhe--ArgLysTyrHisAlaSerAlaIleHisValPro 103
QY      235 GGCCTCAAGCTAAGCGCATTTTCAGATGCAGCCCAAGCGGTTGGAAATCATGTGCACCCCTG 294
Db      104 GlyProGluValAlaArgLeuAlaAspValAlaArgLysAsnHisValTyrLeuValMet 123
QY      295 GGGATGAGTGAACGGGTGCGTGGCACCCCTTACATCAGTCAAGTGTTCATAGGCGATAAT 354
Db      124 GlyAlaIleGlyLysGlyGlyTyrThrLeuTyrCysThrValLeuPhePheSerProGln 143
QY      355 GGTGACACCATTTGGGGCGCGGGAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTTC 414
Db      144 GlyLnPheLeuGlyLysHisArgLysLeuMetProThrSerLeuGluArgCysIleTrp 163
QY      415 GCGCAAGGGAGTATGTTCAATCGCTACGCGTTTTCGAGACGTCGTGTGGAAGCGTGGTGC 474
Db      164 GlyGlnGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGlyAla 183
QY      475 TTATGCTGTGGAGCAGCACCCTTCAACCGCTAACAAATAACGCTTGTATGCACAATAATGA 534
Db      184 AlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaLysGlyIle 203
QY      535 GAGATTCATGTGCG-----GCTTGGCCGAGCTTTAGCCTTAT 573
Db      204 GluLeuTyrCysAlaProThrAlaAspGlySerLysGluTrpGlnSerSerMetLeuHis 223
QY      574 CCTAATCGCGGAAGCCCTGGGCGCTGATGTCATGTAGCGGCGCTCTGAAATCTATGCC 633
Db      224 -----IleAla 225
QY      634 GTTGAAGGCAATGCTTCGTAAGCGTCGTGCGCTGTTTCACAATCCATGATC--- 690
Db      226 IleGluGlyGlyCysPheValLeuSerAlaCysGlnPheCysGlnArgLysHisPhePro 245
QY      691 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCGTGTGCTTCTG 732
Db      246 AspHisProAspTyrLeuPheThrAspTrpTyrAspAspLysGluHisAspSerIleVal 265
QY      733 GCTGGTGTGGACACTCACGTAATAGAGGCGCTGATGGTGTGACTTGTCGCGCCTTT 792
Db      266 SerGlnGlyGly--SerValIleIleSerProLeuGlyGlnValLeuAlaGlyProAsn 284
QY      793 GCCGAAATGAAGAGGTATTTCTTACGCAACCTTGATCCTGGAGTAGCATCCTTGCT 852
Db      285 PheGluSer--GluGlyLeuValThrAlaAspIleAspLeuGlyAspIleAlaArgAla 303
QY      853 AAAATGGCGGACAGCCCTGCTGTCATTATTCGCCGTCGACATTACTGCTGTGTAATA 912
Db      304 LysLeuTyrPheAspSerValGlyTyrTyrSerArgProAspValLeuHisLeuThrVal 323
QY      913 GATCGCAGCCCTAAATTACCGGTA 936
Db      324 AsnGluHisProArgLysSerVal 331
```

```
RESULT 10
S31969
nitriIase (EC 3.5.5.1) [similarity] - Arabidopsis thaliana
N/Alternate names: T10D17.90
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C/Accession: S31969; T49146; T52258
R/Bartling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.
submitted to the EMBL Data Library, September 1992
A/Description: Arabidopsis thaliana nitriIases are encoded by a two-member gene family:
A/Reference number: S31969
A/Accession: S31969
A/Molecule type: mRNA
```



A/Residues: 1-339 <BAR>  
 A/Cross-references: UNIPROT:P32962; UNIPARC:UPI0000001444; EMBL:X68305; NID:g22655; PID  
 R/D/Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; 1  
 submitted to the Protein Sequence Database, April 2000  
 A/Reference number: Z25017  
 A/Accession: T49146  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-339 <DAN>  
 A/Cross-references: UNIPARC:UPI0000001444; EMBL:AL353865; GSPDB:GN00061; ATSP:T10D17.90  
 A/Experimental source: cultivar Columbia; BAC clone T10D17  
 R/artling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 6021-5, 1994  
 A/Title: Molecular characterization of two cloned nitrilases from Arabidopsis thaliana:  
 A/Reference number: Z24514; MUID:8016109; PMID:8016109  
 A/Accession: T52258  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-339 <BA2>  
 A/Cross-references: UNIPARC:UPI0000001444; EMBL:X68305; PIDN:CAA48377.1  
 A/Experimental source: cultivar Landsberg erecta  
 C/Genetics:  
 A/Gene: ATSP:T10D17.90  
 A/Map position: 3  
 A/Introns: 37/1; 97/1; 195/1; 289/1  
 C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
 C/Keywords: hydrolase

Alignment Scores:  
 Pred. No.: 2.54e-38 Length: 339  
 Score: 492.50 Matches: 112  
 Percent Similarity: 51.2% Conservative: 61  
 Best Local Similarity: 33.1% Mismatches: 110  
 Query Match: 27.0% Gaps: 55  
 DB: 2 Indels: 7

US-09-751-299-3 (1-1014) x S31969 (1-339)

QY	13	ATCAAGTCGCGCTGCGTCAAGCCGCCCGATCTACATGATTTGGAGCGCAGGTGAC	72
Db	18	ValArgAlaThrIleValGlnAlaSerThrValTyrAsnAspThrProAlaThrLeuGlu	37
QY	73	AAAAACATTGAGTTGATGGAAGACACGACGTAATAATGCTGCTGATCGCCTTCCG	132
Db	38	LysAlaAsnLysPheIleValGlnAlaAlaSerLysGlySerGluLeuValPhePro	57
QY	133	GAACCTTGATTCAGGCTACCCATGCTTCTTGTGCTTGACTCACAGCATGGCAATG	192
Db	58	GluAlaPheIleGlyTyrPro-----ArgGlyPhe	68
QY	193	CAATT-----GTACGCCAATAC	210
Db	69	ArgPheGlyLeuGlyValGlyValHisAsnGluGlyArgAspGluPheArgLysTyr	88
QY	211	CATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCCGATTTCAGATGCAGCCAAG	270
Db	89	HisAlaSerAlaIleLysValProGlyProGluValGluLysLeuAlaGluLeuAlaGly	108
QY	271	CGGTTGGGAATCATGGTCAACCTGGGGATGAGTGAACGGGTGGTGCCACCTTTACATC	330
Db	109	LysAsnAsnValTyrLeuValMetGlyAlaIleGluLysAspGlyTyrThrLeuTyrCys	128
QY	331	AGTCAGTGGTTCATAGGGCGATAATGCTGACACCATTTGGGGCGCGAAAGTTGAAACCT	390
Db	129	ThrAlaLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHisArgLysLeuMetPro	148
QY	391	ACTTTGTTGAACGCTACTTTGTTGGCGGAAGGGGATGTTGATCGCTAGCGTTTTCGAG	450
Db	149	ThrSerLeuGluArgCysIleTyrGlyGlnGlyAspGlySerThrIleProValTyrAsp	168
QY	451	ACGCTGTTGGAAGCGTGGGTGCTTATGCTGTTGGAGACACCTTCAACCGCTAACAA	510
Db	169	ThrProIleGlyLysLeuGlyAlaAlaIleCysTyrGluAsnArgMetProLeuTyrArg	188

```

QY      511  TACGCTTTGATGCACAAAATGAAGATTCTATTGTGCG----- 549
           |||||
Db      189  ThrAlaLeuTyraIaIySgLyIleGluLeuTyrCySaIaProThrAlaAspGlySerLys 208
QY      550  GCTTGCGCCGAGCTTTAGCCCTTTATCTTAATGCGCGAAAGCCCTGGGCGCTGATGTCAAT 609
           |||
Db      209  GluTrpGlnSerSerMetLeuHis----- 216
QY      610  GTAGCGGCGCTTCGAATCTATGCGCGTTGAAGGGCAATGCTTCGTAAGCGTCGTGCG 669
           |||
Db      217  -----IleAlaIleGluGlyGlyCysPheValIleuSerAlaCysGln 230
QY      670  CTCGTTTCACAATCCATGATC-----GATATGCTTTGTACAGAT-----GAC 711
           ::
Db      231  PheCysLeuArgLysAspPheProAspHisProAspTyrLeuPheThrAspTrpTyrAsp 250
QY      712  GAAAGCATGCGTGTCTTCTTGCGCTGGTGGACACTCACCGTATCATAGGGCGCTGATGCT 771
           ::|||
Db      251  AspLysGluProAspSerIleValSerGlnGlyGlySerValIleIleSerProLeuGly 270
QY      772  GGTGACTTGGTGGCGCGCTCTTGCCGAAATGAAGAGGATTTCTTACCCAACTTGAT 831
           |||
Db      271  GluValLeuAlaGlyProAsnPheGluSer--GluGlyLeuIleThrAlaAspLeuAsp 289
QY      832  CCTGAGATACGCATCCTTGCTAAATGGCGGAGACCCCTGCTGCTCATTAATCCCGTCCC 891
           |||
Db      290  LeuGlyAspValAlaIaArgAlaLysLeuTyrrPheAspSerValGlyHisTyrSerArgPro 309
QY      892  GACATTACTCGCTTGCTTAATAGATCGCAGCCCTAAATTACCGGTAAGTGAATT 945
           |||
Db      310  AspValLeuHisLeuThrValAsnGluHisProLysLysProValThrPheIle 327

RESULT 11
T52259
nitriIase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31-Dec-2004
C;Accession: T52259
R;Hillebrand, H.; Bartling, D.; Weiler, E.W.
Plant Mol. Biol. 36, 89-99, 1998
A;Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitriIases, enz
A;Reference number: Z26007; MUID:98145459; PMID:9484465
A;Accession: T52259
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-339 <HIL>
A;Cross-references: UNIPROT:O04907; UNIPARC:UPI00000A5899; EMBL:Y07648; PIDN:CAA68934.3
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: nit2
A;introns: 37/1; 97/1; 195/1; 289/1
C;Superfamily: nitriIase (carbon-nitrogen hydrolase)
C;Keywords: hydrolase

Alignment Scores:
Pred. No.:      3.16e-38      Length:      339
Score:          491.50      Matches:      112
Percent Similarity: 51.2%      Conservative: 61
Best Local Similarity: 33.1%      Mismatches: 110
Query Match:    26.9%      Indels:      55
DB:             2          Gaps:          7

US-09-751-299-3 (1-1014) x T52259 (1-339)
QY      13  ATCAAGTCGCGCTGCGTGCAGAGCCGCCGATCTACATGATTTGGAGCGGACGGTGAC 72
           :::::
Db      18  ValArgAlaThrIleValGlnAlaSerThrValTyrAsnAspThrProAlaThrLeuGlu 37
QY      73  AAAACCATGAGTTGATGGAAGAGCAGACGTAATATGCTCGTCTGATCGCCTTCCG 132
           |||
Db      38  LysAlaAsnLysPheIleValGluAlaIaSerLysGlySerGluLeuValIaIaPhePro 57

```

QY 133 GAAACTTGATTCAGGCTACCCATGGTTTCTTGCTTGACTACACGACATGGCAATG 192  
|||:::|||||  
Db 58 GluAlaPheIleGlyGlyTyrPro-----ArgGlyPhe 68  
QY 193 CAATT-----GTACGCCAATAC 210  
:::|||  
Db 69 ArgPheGlyLeuGlyValGlyValHisAsnGluGluGlyArgAspGluPheArgLysTyr 88  
QY 211 CATGAGAACTCATGTGAGTTGAGTGGCCCTCAAGCTAAGCCATTTCAGATGCAGCCCAAG 270  
|||:::|||  
Db 89 HisAlaSerAlaIleLysValProGlyProGluValGluLysLeuAlaGluLeuAlaGly 108  
QY 271 CGGTGGGAATCATGCTACCCCTGGGATGAGTGAACGGGTGGTGACCCCTTATCATC 330  
:::|||:::|||  
Db 109 LysAsnAsnValTyrLeuValMetGlyAlaIleGluLysAspGlyTyrThrLeuTyrCys 128  
QY 331 AGTCAGTGTTCATAGGCGATATGTTGACACACCATGGGGCCCGGAAAGTTGAACCT 390  
:::|||  
Db 129 ThrAlaLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHisArgLysLeuMetPro 148  
QY 391 ACTTTGTGAACGTACTTGTTCGGCGAAGGGAGTGTTCATCGCTAGCGGTTTCGAG 450  
|||:::|||  
Db 149 ThrSerLeuGluArgCysIleTyrGlyGlnGlyAspGlySerThrIleProValTyrAsp 168  
QY 451 ACGTCTGTGGAAGCGTGGTGGCTTATGCTGTGGAGCACCTTCAACCGCTAACAAAA 510  
|||:::|||  
Db 169 ThrProIleGlyLysLeuGlyAlaAlaIleCysTyrGluAsnArgMetProLeuTyrArg 188  
QY 511 TACGCTTTGTATGCACAAAATGAAGATTCATTGTGCG----- 549  
|||  
Db 189 ThrAlaLeuTyrAlaLysGlyIleGluLeuTyrCysAlaProThrAlaAspGlySerLys 208  
QY 550 GCTTGCGCCGAGCTTTAGCCTTATCTTAATGCGCGAAAGCCCTGGGCGCTGATGCAAT 609  
|||  
Db 209 GluTyrGlnSerSerMetLeuHis----- 216  
QY 610 GTAGCGGCTCTCGAATGTATGCGCTGAAGGCAATGCTTCTAGTACGCTGTGCG 669  
217 -----IleAlaIleGluGlyCysPheValLeuSerAlaCysGln 230  
QY 670 CTCGTTTACATCCATGATC-----GATATGCTTTGTACAGAT-----GAC 711  
:::|||  
Db 231 PheCysLeuArgLysAspPheProAspHisProAspTyrLeuPheThrAspTyrTyrAsp 250  
QY 712 GAAAGCATGCGTTCCTTCGCTGCTGCTGTGTGACACTCAGTATCATAGGGCTGATGGT 771  
:::|||  
Db 251 AspLysGluProAspSerIleValSerGlnGlySerValIleIleSerProLeuGly 270  
QY 772 GGTGACTTGTGCGCCTCTTGCCGAAATGAAGGGTATTTCTCTACGAAACCTTGAT 831  
|||  
Db 271 GluValLeuAlaGlyProAsnPheGluSer--GluGlyLeuIleThrAlaAspLeuAsp 289  
QY 832 CTTGAGTAGCATCCCTTGCTAAATGGCGGAGACCCCTGCTGTGTCATTATCCCGTCCC 891  
|||  
Db 290 LeuGlyAspValAlaArgAlaLysLeuTyrPheAsp\*\*ValGlyHisTyrSerArgPro 309  
QY 892 GACATTACTCGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATT 945  
|||  
Db 310 AspValLeuHisLeuThrValAsnGluHisProLysLysProValThrPheIle 327

RESULT 12  
T49148  
nitrilase (EC 3.5.5.1) 3 [imported] - Arabidopsis thaliana  
N;Alternate names: protein T10D17.110  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 31-Dec-2004  
C;Accession: T49148; T52261; T52264  
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25017  
A;Accession: T49148  
A;Status: preliminary  
A;Molecule type: DNA

A;Residues: 1-346 <DAN>  
A;Cross-references: UNIPROT:P46010; UNIPARC:UPI000000BDB6; EMBL:ALJ53865; GSPDB:GN00061,  
A;Experimental source: cultivar Columbia; BAC clone T10D17  
R;Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A;Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A;Reference number: Z26007; MUID:98145459; PMID:9484465  
A;Accession: T52261  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-346 <HIL>  
A;Cross-references: UNIPARC:UPI000000BDB6; EMBL:Y07648; PIDN:CAA68936.2  
A;Experimental source: cultivar Columbia  
R;Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A;Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A;Reference number: Z24515; MUID:8022831; PMID:8022831  
A;Accession: T52264  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-346 <BAR>  
A;Cross-references: UNIPARC:UPI000000BDB6; EMBL:U09959; PIDN:AAA19627.1  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: ATSP:T10D17.110; nit3  
A;Map position: 3  
A;Intons: 44/1; 104/1; 202/1; 296/1  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase  
  
Alignment Scores:  
Pred. No.: 3.95e-38 Length: 346  
Score: 490.50 Matches: 115  
Percent Similarity: 55.9% Conservative: 65  
Best Local Similarity: 35.7% Mismatches: 113  
Query Match: 26.9% Indels: 29  
Gaps: 8  
  
US-09-751-299-3 (1-1014) x T49148 (1-346)  
QY 13 ATCAAGTGCCTGCGTGCAAGCCGCCCATCTACATGATTTGGAGCGACGTTGAC 72  
:::|||  
Db 25 ValArgValThrIleValGlnSerSerThrValTyrAsnAspThrProAlaThrLeuAsp 44  
QY 73 AAACCATTTGATTTGGAAGAGCAGCAGTAAATATGCTCTGCTGATCGCCTTCCG 132  
|||  
Db 45 LysAlaGluLysPheIleValGluAlaIleSerLysGlyAlaLysLeuValLeuPhePro 64  
QY 133 GAAACTTGATTCAGGCTACCCA-----TGCTTCTTTGGCTTGAC 174  
|||  
Db 65 GluAlaPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuAlaValGlyValHis 84  
QY 175 TCACACGATGGGCAATGCAATTTGTACGCCAATACCATGAGACTCATTTGAGTTGGAT 234  
:::|||  
Db 85 AsnGluGluGlyArgAspGluPhe---ArgAsnTyrHisAlaSerAlaIleLysValPro 103  
QY 235 GGCCCTCAAGCTAAGCGCATTTCAGATGCAGCCCAAGCGGTTGGGAATCATGTCACCTG 294  
|||  
Db 104 GlyProGluValGluArgLeuAlaGluLeuAlaGlyLysAsnAsnValHisLeuValMet 123  
QY 295 GGGATGATGAACGGGTGGTGCGACCTTTACATCATGTCAGTGGTTGTCATAGCGGATAT 354  
|||  
Db 124 GlyAlaIleGluLysAspGlyTyrThrLeuTyrCysThrAlaLeuPhePheSerProGln 143  
QY 355 GGTGACACCATTTGGGGCCCGGGAAGTTGAACCTACTTTTGTGACGTACTTTGTTG 414  
|||  
Db 144 GlyGlnPheLeuGlyLysHisArgLysValMetProThrSerLeuGluArgCysIleTyr 163  
QY 415 GCGAAGGGGATGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGAAGGCTGGTGCG 474  
|||  
Db 164 GlyGlnGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysIleGlyAla 183  
QY 475 TTATGCTGTGGGAGCACCTTCAACCGCTAACAAATATACGCTTTGTATGCACAAATGAA 534

Db 184 AlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaLysGlyIle 203  
QY 535 GAGATTCATGTGCGGCTTGCGCGAGCTTTAGCCTTATCCTAATAGCGGCGAAAGCCCTG 594  
Db 204 GluIleTyrCysAlaProThrAlaAspTyrSerLeu----- 215  
QY 595 GGGCCTGATGTCAATGTAGCGGCTCTCGAATCTAT---GCCGTGAAGGCAATGCTTC 651  
Db 216 -----GluTrpGlnAlaSerMetIleHisIleAlaValGluGlyCysPhe 231  
QY 652 GTACTAGCGTCGTGCGCTCGTTTCAACAATCCATGATC-----GATATGCTT 699  
Db 232 ValLeuSerAlaHisGlnPheCysLysArgArgGluPheProGluHisProAspTyrLeu 251  
QY 700 -----TGTAAGATGACGAAAGCATGCGCTTGCTTGCGCTGCTGAGCACTCA 750  
Db 252 PheAsnAspIleValAspThrLysGluHisAspProThrValSerGlyGly---Ser 270  
QY 751 CGTATCATAGGGCTGATGGTGGTGACTTGTCGCGCCTCTTGCCGAAATGAGAGGGT 810  
Db 271 ValIleIleSerProLeuGlyLysValLeuAlaGlyProAsnTyrGluSer---GluGly 289  
QY 811 ATTCTCTACGCAAACTTGATCCTGGAGTACGCATCCTTGCTAAATGCGCGAGACCCCT 870  
Db 290 LeuValThrAlaAspLeuAspLeuGlyAspIleAlaArgAlaLysLeuTyrPheAspVal 309  
QY 871 GCTGTCATTATTCCTCCCGACATTACTCGCTTGCTAATAGATCGCAGCCCTAAATTA 930  
Db 310 ValGlyHisTyrSerLysProAspIlePheAsnLeuThrValAsnGluHisProLysLys 329  
QY 931 CCGGTA 936  
Db 330 ProVal 331

RESULT 13  
T52262  
nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52262  
R/Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A/Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A/Reference number: Z24515; MUID:8022831; PMID:8022831  
A/Accession: T52262  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-339 <BAR>  
A/Cross-references: UNIPROT:P32962; UNIPARC:UPI000016DA4C; EMBL:U38845; PIDN:AAB05220.1  
A/Experimental source: cultivar Columbia  
C/Genetics:  
A/Gene: NIT2  
A/Map position: 3  
A/Introns: 37/2; 97/1; 195/1; 289/1  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Alignment Scores:  
Pred. No.: 4.89e-38 Length: 339  
Score: 489.50 Matches: 112  
Percent Similarity: 50.9% Conservative: 60  
Best Local Similarity: 33.1% Mismatches: 111  
Query Match: 26.8% Indels: 55  
DB: 2 Gaps: 7

US-09-751-299-3 (1-1014) x T52262 (1-339)  
QY 13 ATCAAGTCGCTGCGTGAAGCGCCCGATCTACATGATTGGAGGCGAGGTGAC 72  
Db 18 ValArgAlaThrIleValGlnIleAspThrValTyrAsnAspThrProAlaThrLeuGly 37  
QY 73 AAAACCATGTAGTTGATGGAAGAAGACACGTAATAATGCTGCTGATCGCCTTCCG 132

Db 38 LysAlaAsnLysPheIleValGluAlaAlaThrLysGlySerGluLeuValPhePro 57  
QY 133 GAAACTTGATCCAGGCTACCCCATGGTTTCTTGCTTGACTGACTACAGCATGGCAATG 192  
Db 58 GluAlaPheIleGlyGlyTyrPro-----ArgGlyPhe 68  
QY 193 CAATTT-----GTAGCCCAATAC 210  
Db 69 ArgPheGlyLeuGlyValGlyValHisAsnGluGluGlyArgAspGluPheArgLysTyr 88  
QY 211 CATGAGAACTGATTTGAGTTGATGGCCCTCAAGCTTAAGCGCATTTGAGTCAGCCAAAG 270  
Db 89 HisAlaSerAlaIleLysValProGlyProGluValGluLysLeuAlaGluLeuAlaGly 108  
QY 271 CGGTGGGAATCATGTGCACCCCTGGGATGAGTGAACGGGTGGTGACCTTTACATC 330  
Db 109 LysAsnAsnValTyrLeuValMetGlyAlaIleGluLysAspGlyTyrThrLeuTyrCys 128  
QY 331 AGTCAGTGTTCAATAGCGGATATGTTGACACCATTTGGGGCCCGGAAAGTTGAAACCT 390  
Db 129 ThrAlaLeuPhePheSerProGlnGlyGlnPheLeuGlyLysHisArgLysLeuMetPro 148  
QY 391 ACTTTGTTGAACGTACTTTGTTCGGCGGAGGGAGTGGTTCATCGCTAGCGGTTTTCGAG 450  
Db 149 ThrSerLeuGluArgCysIleTyrGlyGlnGlyAspGlySerThrIleProValTyrAsp 168  
QY 451 ACGTCTGTTGAAGGCTGGGTGGCTTATGCTGTGGAGACACCTTCAACCGCTAACAAA 510  
Db 169 ThrProIleGlyLysLeuGlyAlaAlaIleCysTrpGluAsnArgMetProLeuTyrArg 188  
QY 511 TACGCTTTGATGACAAATGAAGAGATTCATGTGCG----- 549  
Db 189 ThrAlaLeuTyrAlaLysGlyIleGluLeuTyrCysAlaProThrAlaAspGlySerLys 208  
QY 550 GCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCTGATGTCAAT 609  
Db 209 GluTrpGlnSerSerMetLeuHis----- 216  
QY 610 GTAGCGGCTCTGCAATCTATGCCGTTGAAGGCAATGCTTGACTAGCGTGTGCG 669  
Db 217 -----IleAlaIleGluGlyGlyCysPheValLeuSerAlaCysGln 230  
QY 670 CTCGTTTCACAATCCATGATC-----GATATGCTTTGTACAGAT-----GAC 711  
Db 231 PheCysLeuArgLysAspPheProAspHisProAspTyrLeuPheThrAspTrpTyrAsp 250  
QY 712 GAAAGCATGCGTTCCTTGCTGCTGGCTGGTGGACACTCACGTATCATAGGCGCTGATGCT 771  
Db 251 AspLysGluProAspSerIleValSerGlnGlySerValIleIleSerProLeuGly 270  
QY 772 GGTGACTTGGTCGCGCTCTTGCCGAAATGAAGGGTATTCTTACGCAACTTGAT 831  
Db 271 GlnValLeuAlaGlyProAsnPheGluSer--GluGlyLeuIleThrAlaAspLeuAsp 289  
QY 832 CTTGAGTAGCGCATCCTTGCTAAATGGCGGAGACCCCTGCTGTCATTATTCGCTCC 891  
Db 290 LeuGlyAspValAlaArgAlaLysLeuTyrPheAspSerValGlyHisTyrSerArgPro 309  
RESULT 14  
T27679  
probable nitrilase (EC 3.5.5.1) ZK1058.6 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Dec-2004  
C/Accession: T27679  
R/Mortimore, B.  
Submitted to the EMBL Data Library, August 1994  
A/Reference number: Z20403  
A/Accession: T27679







```

Db      171 IleIleTrrGlyPheGlyAspGlySerThrIleProValTyrAspThrProLeuGlyLys 190
QY      466 CTGGTGCTTATGCTGTGGAGACCTTCAACCGCTAACAAAATACGCTTGTATGCA 525
      :::||| ||| ||| ||| ||| ::: ||| ||| ||| |||
Db      191 IleGlyAlaLeuIleCysTrrGluAsnLysMetProLeuLeuArgThrAlaLeuTyrGly 210
QY      526 CAAAATGAGAGATTCAATTGTGCGCTTGCGCGAGCTTTAGCCTTTATCTTAATGCGGCG 585
      ::: ||| ||| ::: ||| |||
Db      211 LysGlyIleGluIleTyrCysAla-----ProThrAlaAsp 222
QY      586 AAAGCCCTGGGGCCTGATGTCATGATGACGGCCTCTCGAATCTATGCCGTTGAAGGCAA 645
      ::: ||| ||| ::: ||| |||
Db      223 SerArg-----GlnValTrrGlnAlaSerMetThrHisIleAlaLeuGluGly 239
QY      646 TGCTTCGTAAGCGTGTGCGCTCGCTTCAACAATCC----- 684
      ||| ||| ||| ||| ::: ||| |||
Db      240 CysPheValLeuSerAlaAsnGlnPheCysArgArgLysAspTyrProProProProGlu 259
QY      685 ---ATGATCGATATGCTTTGTACAGATGACGAAAAGCATGCGCTTGTGCTGCTGCTGCT 741
      ::: ||| ||| ::: ||| |||
Db      260 TyrValPheThrGlyLeuGlyGluGluProSerProAspThrValValCysProGlyGly 279
QY      742 GGACACTCACGTATGATGAGGCGCTGATGGTGTGACTTGGTGGCGCTTGGCGAAAT 801
      ::: ||| ||| ||| ||| ||| ||| ||| |||
Db      280 -----SerValIleIleSerProSerGlyGluValLeuAlaGlyProAsnTyrGluGly 297
QY      802 GAAGAGGTATTTCTTACGCAAACTTGATCCTGAGTACGCATCTTGCTAAATGCGG 861
      ::: ||| ::: ||| ||| ||| ||| ||| |||
Db      298 ---GluAlaLeuIleThrAlaAspLeuAspLeuGlyGluIleValArgAlaLysPheAsp 316
QY      862 GCAGACCTGCTGCTCATTTATTCCTCCGTCGACATTAATGCTAATAGATCGCAGC 921
      ::: ||| ||| ||| ||| ::: ||| ||| ||| |||
Db      317 PheAspValValGlyHisTyrAlaArgProGluValLeuSerLeuValValAsnAspGln 336
QY      922 CCTAATTACCGGTA 936
      ||| ||| ||| ||| |||
Db      337 ProHisLeuProVal 341

```

Search completed: April 27, 2006, 01:38:10  
 Job time : 45.981 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:07:06 ; Search time 49.8858 Seconds  
(without alignments)  
2868.171 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1826  
Sequence: 1 atgaaagaagctatcaagt.....cgcaactcgaagaatttga 1014

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090034\_18631/app\_query.fasta\_1  
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US09751299 @CGN\_1\_1\_580 @runat\_26042006\_090034\_18631 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_05.80: \*  
1: uniProt\_sprot: \*  
2: uniProt\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765	96.7	337	2	Q6RWE0_9ZZZZ	Q6rwe0 uncultured
2	1343	73.5	337	2	Q6RWQ3_9ZZZZ	Q6rww3 uncultured
3	1343	73.5	337	2	Q6RWQ7_9ZZZZ	Q6rww7 uncultured
4	1187	65.0	336	2	Q500U1_PSESY	Q500u1 pseudomonas
5	1184	64.8	347	2	Q88B32_PSESM	Q88b32 pseudomonas
6	1139	62.4	333	2	Q6RWL0_9ZZZZ	Q6rwl0 uncultured
7	1136	62.2	332	2	Q6RWJ0_9ZZZZ	Q6rwj0 uncultured
8	1081	59.2	345	2	Q6RWS5_9ZZZZ	Q6rws5 uncultured
9	1051	57.6	334	2	Q6RWE3_9ZZZZ	Q6rwe3 uncultured
10	1042	57.1	353	2	Q6RWF9_9ZZZZ	Q6rwf9 uncultured
11	1039	56.9	348	2	Q6RWS2_9ZZZZ	Q6rws2 uncultured
12	1037	56.8	336	2	Q6RWF6_9ZZZZ	Q6rwf6 uncultured
13	1037	56.8	353	2	Q6RWF0_9ZZZZ	Q6rwf0 uncultured
14	1033	56.6	337	2	Q6RWS3_9ZZZZ	Q6rws3 uncultured
15	1033	56.6	338	2	Q6RWB4_9ZZZZ	Q6rwb4 uncultured
16	1028.5	56.3	358	2	Q6RWI0_9ZZZZ	Q6rwi0 uncultured

17	1022.5	56.0	337	2	Q6RWN8_9ZZZZ	Q6rwn8 uncultured
18	1021.5	55.9	337	2	Q6RWN9_9ZZZZ	Q6rwn9 uncultured
19	1020.5	55.9	354	2	Q6RWK6_9ZZZZ	Q6rwk6 uncultured
20	1019.5	55.8	338	2	Q6RWG6_9ZZZZ	Q6rwg6 uncultured
21	1018	55.8	338	2	Q6RWN7_9ZZZZ	Q6rwn7 uncultured
22	1008	55.2	337	2	Q6RWI7_9ZZZZ	Q6rwi7 uncultured
23	1008	55.2	346	2	Q6RWK5_9ZZZZ	Q6rwk5 uncultured
24	999.5	54.7	338	2	Q6RWN2_9ZZZZ	Q6rwn2 uncultured
25	999	54.7	338	2	Q6RWR6_9ZZZZ	Q6rwr6 uncultured
26	997.5	54.6	336	2	Q5YUM5_NOCFA	Q5yum5 nocardia fa
27	994.5	54.5	338	2	Q6RWQ4_9ZZZZ	Q6rww4 uncultured
28	993	54.4	338	2	Q6RWH1_9ZZZZ	Q6rwh1 uncultured
29	980.5	53.7	327	2	Q6RWH9_9ZZZZ	Q6rwh9 uncultured
30	976	53.5	352	2	Q6RWH0_9ZZZZ	Q6rwh0 uncultured
31	973.5	53.3	336	2	Q6RWJ5_9ZZZZ	Q6rwj5 uncultured
32	971.5	53.2	341	2	Q6RWR4_9ZZZZ	Q6rwr4 uncultured
33	937.5	51.3	335	2	Q6RWF7_9ZZZZ	Q6rwf7 uncultured
34	936	51.3	331	2	Q706Q8_PSEPU	Q706q8 pseudomonas
35	929.5	50.9	337	2	Q6RWM3_9ZZZZ	Q6rwm3 uncultured
36	924	50.6	350	2	Q5EG61_PSEFL	Q5eg61 pseudomonas
37	919.5	50.4	334	2	Q6RWG9_9ZZZZ	Q6rww9 uncultured
38	918.5	50.3	334	2	Q89GE3_BRAJA	Q89ge3 bradyrhizob
39	914.5	50.1	337	2	Q6RWN9_9ZZZZ	Q6rwn9 uncultured
40	906	49.6	345	2	Q6RWN2_9ZZZZ	Q6rwn2 uncultured
41	898.5	49.2	349	2	Q6RWS0_9ZZZZ	Q6rws0 uncultured
42	881	48.2	346	2	Q6RWM1_9ZZZZ	Q6rwm1 uncultured
43	836.5	45.8	365	1	NRL2_RHORH	Q03217 rhodococcus
44	819	44.9	356	1	NRLA_ALCPA	P20960 alcaligenes
45	800	43.8	383	1	NRL1_RHORH	Q02068 rhodococcus

ALIGNMENTS

RESULT 1  
Q6RWE0\_9ZZZZ PRELIMINARY; PRT; 337 AA.  
ID Q6RWE0;  
AC Q6RWE0;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nitriylase (EC 3.5.5.7).  
GN ORFNames=BD5220;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriylase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487562; AAR97509.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitriylase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR00132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 337 AA; 36783 MW; 25D2DE1E820983B2 CRC64;

Alignment Scores:  
Pred. No.: 3.08e-149  
Score: 1765.00  
Length: 337  
Percent Similarity: 100.0%  
Conservative: 337  
Best Local Similarity: 100.0%  
Mismatches: 0  
Query Match: 96.7%  
Indels: 0  
DB: 2  
Gaps: 0



QY 481 TGTGGAGACCTTCACCGCTAACAAATACGCTTGTATGACAAATGAAGATT 540  
Db 161 CystTrpGluHisLeuGlnProLeuSerIleTyrAlaLeuTyrAlaGlnAsnGluGlu 180  
QY 541 CATTTGGCGCTTGCGCGAGCTTTAGCTTTATCCTAATGCGGCGAAAGCCCTGGGCGCT 600  
Db 181 HispheaAlaIleTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGAGCGGCTCTGCAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGluGlnCysPheValLeuAla 220  
QY 661 TCGTGGCGCTCGTTTCACAATCCATGATGATGCTTGTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerIleHis 240  
QY 721 GCGTTGCTTCTGGCTGTGTGAGACCTCACGTCATCATAGGGCTGATGGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGGCGCTCTTGCCGAAATGAAGGGTATTCTCTACGCAACCTGATCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCATTAATCCGTCGCGACATTAAT 900  
Db 281 ArgIleTyrAlaIleThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAGCGTCTGAGACGGGTGCGCACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 3  
Q6RMQ7\_9ZZZZ PRELIMINARY; PRT; 337 AA.  
ID Q6RMQ7\_9ZZZZ  
AC Q6RMQ7;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE NitriLase (EC 3.5.5.7).  
GN ORFNames=BD5274;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitriLase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487445; AAR97392.1; -; Genomic DNA.  
DR GO: GO:0018762; F:aliphatic nitriLase activity; IEA.  
DR GO: GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR000132; N/CN\_hydrtase.  
DR InterPro: IPR003010; NtIse/CNhydase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS50263; CN\_HYDROLASE; 1.  
DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 337 AA; 36769 MW; 940990B5AB69582E CRC64;

Alignment Scores: 2.18e-111 Length: 337  
Pred. No.:

Score: 1343.00 Matches: 253  
Percent Similarity: 84.2% Conservative: 29  
Best Local Similarity: 75.5% Mismatches: 47  
Query Match: 73.5% Indels: 6  
DB: 2 Gaps: 1  
US-09-751-299-3 (1-1014) x Q6RMQ7\_9ZZZZ (1-337)

QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTCAAGCCCGCCGATCTACATGATTTGGAG 60  
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGGTGACAAACCATTGAGTTGATGGAAGACGACGCTAATAATGCTCGTTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATGACCTTTCCGAAACCTTGGATTCCAGGCTACCCATGCTTCTTGGCTTGAATGACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTGTTCAGCCCATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
Db 61 AlaTrpGlyMetGlnPheValAlaArgTrpHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGAAATCATGTGCACCCCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGCGTGCGACCTTTACATCAGTCAGTGGTTCATAGCGGATTAATGTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGCAAGTTGAACCTACTTTGTTGAACGTACTTTGTTGCGCGAA 420  
Db 121 ThrAlaGlyLeuAlaArgAlaGlyLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGTTTCATCGCTTACCGGCTTTTCGAGACGTCGTTGGAAGGCTGGTGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGGAGACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGATT 540  
Db 161 CystTrpGluHisLeuGlnProLeuSerIleTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATTTGGCGCTTGCGCGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGCGCT 600  
Db 181 HispheaAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGATGAGCGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTGACTAGCG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGluGlnCysPheValLeuAla 220  
QY 661 TCGTGGCGCTCGTTTCACAATCCATGATGATGCTTGTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerIleHis 240  
QY 721 GCGTTGCTTCTGGCTGTGTGAGACCTCACGTCATCATAGGGCTGATGGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGGCGCTCTTGCGCGAAATGAAGGGTATTCTCTACGCAACCTGATCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCATTAATCCGTCGCGACATTAAT 900  
Db 281 ArgIleTyrAlaIleThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAGCGTCTGAGACGGGTGCGCACTCGAA 1005



Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 4

Q500U1\_PSESY

ID Q500U1\_PSESY PRELIMINARY; PRT; 336 AA.

AC Q500U1\_

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, last annotation update)

DE Alphatic nitrilase (EC 3.5.5.7).

GN ORFNames=Peyr\_0007;

OS Pseudomonas syringae pv. syringae B728a.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=205918;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RG DOE Joint Genome Institute;

RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,

RA Nolan M., Goltsman E., Thiel J., Malfatti S., Lapidus A., Deter J.C.,

RA Land M., Richardson P.M., Kyriides N.C., Ivanova N.;

RT "Comparison of two complete genome sequences of Pseudomonas syringae

RT pv. syringae B728a and pv. tomato DC3000.";

RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).

RN [2]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RA Loper J.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

RN [3]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=B728a;

RA Feil H., Feil W.S., Lindow S.E.;

RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.

DR EMBL; CP000075; AAY35081.1; -; Genomic\_DNA.

DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

KW Hydrolase.

SQ SEQUENCE 336 AA; 36552 MW; EC3DF2A762E37729 CRC64;

Alignment Scores:

Pred. No.: 2.14e-97 Length: 336

Score: 1187.00 Matches: 218

Percent Similarity: 81.0% Conservative: 37

Best Local Similarity: 69.2% Mismatches: 60

Query Match: 65.0% Indels: 0

DB: 2 Gaps: 0

US-09-751-299-3 (1-1014) x Q500U1\_PSESY (1-336)

QY 1 ATGAAGAAGCTATCAAGTCGCCCTGCGTGAAGCCGCCCGATCTACATGATTGGAG 60

Db 1 MetLysGluProLeuLysValAlaCysValGlnAlaAlaProValPheLeuAspLeu 20

QY 61 GCGACGGTGGACAAACCATTGAGTTGATGGAAGACAGCAGCACTAATATGCTGCTG 120

Db 21 AlaThrValAspLysThrIleThrLeuMetGluGlnAlaAlaAlaGlyAlaGlyLeu 40

QY 121 ATGCCCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180

Db 41 IleAlaPheProGluThrIleProGlyTyrProIlePheLeuTyrPheLeuAspAlaPro 60

QY 181 GCATGGGCAATGCAATTGTAGCGCCATATCCATGAGAACTCATTTGAGTTGGAGCCCT 240

Db 61 AlaTyrAsnMetProLeuValGlnArgTyrHisGlnGlnSerLeuValLeuAspSerVal 80

QY 241 CAAGCTAAGCGCATTTGAGATGACGCCAAGCGGTGGGAATCATGGTCACCCCTGGGATG 300

Db 81 GlnAlaArgArgIleSerAspAlaAlaArgHisLeuGlyLeuTyrValValLeuGlyTyr 100

QY 301 AGTGAACGGGTGCGTGGACACCTTTTACATCAGTCAGTGGTTTCATAGCGGATAATGCTGAC 360

Db 101 SerGluArgAsnLysAlaSerLeuTyrIleGlyGlnTyrIleIleAspAspHisGlyGlu 120

QY 361 ACCATTGGGCGCGCGGAAAGTTGAAACTTCTTTGTTGAACGTACTTGTTCGGCGAA 420

Db 121 ThrValGlyValArgArgLysLeuLysAlaThrHisValGluArgHisMetPheGlyGlu 140

QY 421 GGGATGATTCAATCGCTAGCGGTTTTCGAGACGTCGTGTTGGAAGCTGGGTGCTTATGC 480

Db 141 GlyAspGlyAlaSerLeuArgThrPheGluThrProValGlyValLeuGlyAlaLeuGly 160

QY 481 TGTGGAGACCTTCAACCGCTAACAATAACGCTTTGATGCACAAATGAAGAGATT 540

Db 161 CysTyrGluHisLeuGlnProLeuSerIleTyrAlaMetTyrAlaGlnAsnGluGlnIle 180

QY 541 CATGTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCCT 600

Db 181 HisValAlaAlaTyrProSerPheSerLeuTyrArgAsnAlaThrSerAlaLeuGlyPro 200

QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCGCAATGCTTCGTACTAGCG 660

Db 201 GluValAsnThrAlaAlaAlaSerArgValTyrAlaAlaGluGlyGlnCysPheValLeuAla 220

QY 661 TCGTGTGCGCTGTTTCAACAATCCATGCATGATATGCTTTGTACAGATGACGAAAGCAT 720

Db 221 ProCysAlaIleValIleSerProGluMetIleGluMetLeuCysAspSerAspAlaLysArg 240

QY 721 GCGTTGCTTTCGCTGCTGTGTGACACTACGATCATAGGCGCTGATGCTGACTTG 780

Db 241 SerLeuLeuGlnAlaGlyGlyHisAlaArgIlePheGlyProAspGlySerAspLeu 260

QY 781 GTCGCGCTCTTCCGAAATGAAGGGTATTCTTACGCAAACTTGATCCTGGAGTA 840

Db 261 AlaThrProLeuGlyGluHisGluGlyLeuLeuTyrAlaThrLeuAspProAlaAla 280

QY 841 CGCATCCTTGTAAATGCGCGAGACCTGCTGTCATTATCCCGTCCGACATTACT 900

Db 281 LeuThrLeuAlaLysValAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300

QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATT 945

Db 301 ArgLeuMetPheAsnProAsnProThrProCysValValAspLeu 315

RESULT 5

Q88B32\_PSESM

ID Q88B32\_PSESM PRELIMINARY; PRT; 347 AA.

AC Q88B32\_

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)

DE Nitrilase, putative.

GN OrderedLocusNames=PSPT00189;

OS Pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=323;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=DC3000;

RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;

RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,

RA Gwin M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,

RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,

RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,

RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,

RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,

RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,

RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,

RA Bender C.L., White O., Fraser C.M., Collier A.;

RT "The complete genome sequence of the Arabidopsis and tomato pathogen

RT Pseudomonas syringae pv. tomato DC3000.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).

DR EMBL; AB016853; AA053742.1; -; Genomic\_DNA.

TIGR; PSPT00189; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0016810; F:nitrogen compound metabolism; IEA.  
DR GO; GO:006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS0263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 347 AA; 37564 MW; B671F783D7411308 CRC64;

Alignment Scores:  
Pred. No.: 4e-97 Length: 347  
Score: 1184.00 Matches: 218  
Percent Similarity: 80.4% Conservative: 37  
Best Local Similarity: 68.8% Mismatches: 62  
Query Match: 64.8% Indels: 0  
DB: 2 Gaps: 0

US-09-751-299-3 (1-1014) x Q88B32\_PSESM (1-347)

QY 1 ATGAAGAAGCTATCAAGTCCGCTGCTGCAAGCCGCCGATCTACATGATTGGAG 60  
Db MetlysAlaProleuValAlaCysValGlnAlaAlaPrometPheLeuAsnLeuAsp 31  
QY 61 GCGACGGTGGACAAACCATGAGTTGATGGAGAAGCAGCAGTAATATGCTCGTCTG 120  
Db AlaThrValAspLysThrIleAlaLeuIleGlnAlaAlaAlaIleAlaGlyLeu 51  
QY 121 ATGCGCTTTCGGAACTTGGATTCCAGGCTACCCCATGTTTCTTGGCTTGACTCACA 180  
Db IleAlaPheProGluThrTrpIleProGlyTrpOtrPheLeuTrpLeuAspAlaPro 71  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAAGTAATGAGTTGATGGCCCT 240  
Db AlaTrpAsnMetProleuValIleArgTyrHisGlnGlnSerLeuValLeuAspSera 91  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCCAAGCGGTTGGGAATCATGTCACCCCTGGGATG 300  
Db GlnAlaArgArgIleSerAspAlaAlaArgGlnHisGlyIleTyrValValLeuGlyTyr 111  
QY 301 AGTGAACGGGTGGTGGACCCCTTACATCAGTCAAGTGGTTCATAGGCGATTAATGGTAC 360  
Db SerGluArgSerLysAlaSerLeuTyrIleGlyGlnTrpIleIleAspAspGlnGlyGlu 131  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
Db ThrValGlyThrArgArgLysLeuValAlaThrHisValGluArgThrLeuPheGlyGlu 151  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGCGTGGTGGCTTATGC 480  
Db GlyAspGlyAlaSerLeuArgProPheGluThrProValGlyValLeuGlyAlaLeuCys 171  
QY 481 TGTGGGAGCACCTTCAACCGCTTACAACAATACGCTTGTATGCACAATAATGAGAGATT 540  
Db CysTrpGluHisLeuGlnProleuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnIle 191  
QY 541 CATGTGCGGCTTGGCCGAGCTTTAAGCTTATCCTAATGCCGGAAGCCCTGGGCGCT 600  
Db HisValAlaAlaTrpProSerPheSerLeuTyrGlnAsnAlaThrArgAlaLeuGlyPro 211  
QY 601 GATGTCATGATGACGGGCTCTCGAATCTATGCCGTTGAAGGCGCAATGCTTCTGTAAGCG 660  
Db GluValAsnThrAlaAlaSerArgValTyrAlaAlaGluGlyGlnCysPheValIleAla 231  
QY 661 TCGTGTGCGCTGTTTCAATTCATGATGATGATGCTTGTATCAGATGACGAAAGCAT 720  
Db ProCysAlaValAlaSerProGluMetIleGluMetLeuCysAspSerAspAlaLysHis 251  
QY 721 GCGTGTGCTTGTGCTGTGTGAGACACTCAGCTATGAGGCGCTGATGCTGACTTG 780  
Db SerLeuLeuGlnAlaGlyGlyGlyHisAlaArgIlePheGlyProAspGlySerAspLeu 271

QY 781 GTGCGCCTCTTGCCGAAATGAAAGGGTATTCTTACGCAAACTTGATCCTGAGTA 840  
Db AlaThrProleuGlyGlnHisGluGlyLeuLeuTyrAlaThrLeuAspProAlaAla 291  
QY 841 CGCATCCTTGTAAATGGCGGAGACCCCTGCTGTCATTATTCCTCCGACATTA 900  
Db LeuIlePheAlaLysValAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 311  
QY 901 CGCTTGTCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATTGAAGGT 951  
Db ArgLeuMetPheAsnProAsnProAsnProCysValValGluLeuProGly 328

RESULT 6  
Q6RWL0\_9ZZZZ  
ID Q6RWL0\_9ZZZZ PRELIMINARY; PRT; 333 AA.  
AC Q6RWL0;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nitriase (EC 3.5.5.7).  
GN ORFNames=BD7615;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxId=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Roberton D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitriase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487492; AAR97439.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitriase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS0263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
KW Hydrolyase.  
SQ SEQUENCE 333 AA; 36342 MW; B0A9B825FD0D7403 CRC64;

Alignment Scores:  
Pred. No.: 4.31e-93 Length: 333  
Score: 1139.00 Matches: 212  
Percent Similarity: 79.0% Conservative: 36  
Best Local Similarity: 67.5% Mismatches: 66  
Query Match: 62.4% Indels: 0  
DB: 2 Gaps: 0

US-09-751-299-3 (1-1014) x Q6RWL0\_9ZZZZ (1-333)

QY 1 ATGAAGAAGCTATCAAGTCCGCTGCTGCAAGCCGCCGATCTACATGATTGGAG 60  
Db MetLysThrThrValThrValAlaCysValGlnAlaAlaProValPheMetAspLeuGlu 21  
QY 61 GCGACGGTGGACAAACCATGAGTTGATGGAAGCAGCAGCAGTAATATGCTCGCTG 120  
Db GlyThrIleAspLysThrIleThrLeuIleSerGluAlaAlaGlnLysGlyAlaGluLeu 41  
QY 121 ATGCGCTTTCGGAACTTGGATTCCAGGCTACCCATGTTTCTTGGCTTGACTCACA 180  
Db IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAsnSerPro 61  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAAGTAATGAGTTGAGCCCT 240  
Db AlaThrAsnMetProleuValTyrGlnTyrHisGlnAsnSerLeuValLeuAspSera 81

QY 241 CAAGCTAAGCGCATTTTCAGATGACGCAAGCGGTGGGAATCATGTGACCCCTGGGATG 300  
|||||  
Db 82 GlnAlaIysArgIleAlaAspAlaIaGlnGlnAsnIleThrValValLeuGlyPhe 101  
QY 301 AGTGAACGGGTGGTGGCAACCCCTTACATCAGTGGTTCATAGCGGATAATGCTGAC 360  
|||||  
Db 102 SerGluArgAspHisIsgIysSerLeuTyrlIeSerGlnTrpLeuIleGlySerAspGlyGlu 121  
QY 361 ACCATTGGGGCGCGCAAGTTGAACCTACTTTGTGTAACGCTACTTTGTTCGGCGAA 420  
|||||  
Db 122 ThrIleGlyIleArgArgIysLeuLysAlaThrHisValGluArgThrLeuPheGlyGlu 141  
QY 421 GGGGATGGTTCATCCGCTAGCCGTTTTCGAGACGCTCTGTGGAAGCGTGGGCTTATGC 480  
|||||  
Db 142 SerAspGlySerSerLeuThrThrTrpGluThrProLeuGlyAsnValGlyAlaLeuCys 161  
QY 481 TGTGTGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
|||||  
Db 162 CysTrpGluHisLeuGlnProLeuSerArgIlyrAlaMetTyrSerGlnHisGluGluIle 181  
QY 541 CATTTGCGCGCTTGGCCGAGCTTTAGCCCTTATCTAATGCGGCGAAAGCCCTGGGCCCT 600  
|||||  
Db 182 HisIleAlaIaIaTrpProSerPheSerLeuTyThrSerAlaThrAlaIaLeuGlyPro 201  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCGGTGAAGGCGCAATGCTTCTACTAGCG 660  
|||||  
Db 202 AspValAsnThrAlaIaIaSerArgLeuTyThrAlaIaGluGlyGlnCysPheValLeuAla 221  
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATCGATGATGCTTGTGTACAGATGACGAAAGCAT 720  
|||||  
Db 222 ProCysAlaValValSerAspGluMetIleAspLeuLeuCysProAspAspArgArg 241  
QY 721 GCGTTGCTTCTGCTGTGGTGGAGACACTACGATCATAGGGCTGATGGTGTGACTTG 780  
|||||  
Db 242 AlaLeuLeuSerAlaGlyGlyGlyHisAlaArgIleTyrglyProAspGlyArgGluLeu 261  
QY 781 GTCGGCGCTTGTCCGAAATGAAGAGGATATCTCTACGCAACCTGATCCTCGAGTA 840  
|||||  
Db 262 ValThrProLeuGlyGluAsnGluGlyLeuLeuIleAlaGluLeuAspSerAlaAla 281  
QY 841 CGCATCCTTGTCTAAATGGCGGAGACCCCTGCTGTCTATTATCCCGTCCCGACATTACT 900  
|||||  
Db 282 IleThrPheAlaIysLeuAlaIaAspProValGlyHisTyrSerArgProAspValThr 301  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAA 942  
|||||  
Db 302 ArgLeuLeuPheAsnProSerAlaAsnLysThrValIleLys 315  
RESULT 7  
Q6RMJ0\_9ZZZZ PRELIMINARY; PRT; 332 AA.  
ID Q6RMJ0\_9ZZZZ  
AC Q6RMJ0\_9ZZZZ  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5168;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487512; AAR97459.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.

DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; NtIse/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS0263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
KW Hydrolyase.  
SQ SEQUENCE 332 AA; 36214 MW; 2E77B8470F426184 CRC64;  
Alignment Scores:  
Pred. No.: 8e-93 Length: 332  
Score: 1136.00 Matches: 211  
Percent Similarity: 79.0% Conservative: 37  
Best Local Similarity: 67.2% Mismatches: 66  
Query Match: 62.2% Indels: 0  
DB: 2 Gaps: 0  
US-09-751-299-3 (1-1014) x Q6RMJ0\_9ZZZZ (1-332)  
QY 1 ATGAAGAAGCTATCAAGTCCGCTGCGTGAAGCCCGCCGATCTACATGATTGGAG 60  
|||||  
Db 1 MetLysThrThrValThrValAlaCysValGlnAlaIaIaProValPheMetAspLeuGlu 20  
QY 61 GCGAGGTGGACAAACCATGTAGTGTGTAAGAAGACGACGACGTAATATGCTCGTCTG 120  
|||||  
Db 21 GlyThrValAspLysThrIleThrLeuIleSerGluAlaIaGlnLysGlyAlaGluLeu 40  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180  
|||||  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAsnSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGTTGGATGGCCCT 240  
|||||  
Db 61 AlaThrAsnMetProLeuValTyrglnTyHisGlnAsnSerLeuValLeuAspSerThr 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCAAGCGGTGGGAATCATGCTGACCCCTGGGATG 300  
|||||  
Db 81 GlnAlaIysArgIleAlaAspAlaIaArgGlnAsnAsnIleThrValAlaLeuGlyPhe 100  
QY 301 AGTGAACGGGTGGTGGCACCCCTTTCATCAGTCACTGTTTCATAGCGGATAATGCTGAC 360  
|||||  
Db 101 SerGluArgAspHisIsgIysSerLeuTyrlIeAlaGlnTrpLeuIleGlySerAspGlyGlu 120  
QY 361 ACCATTGGGGCGCGCAAGTTGAACCTACTTTGTGTAACGTAATGCTTGTTCGGCGAA 420  
|||||  
Db 121 ThrIleGlyIleArgArgIysLeuLysAlaThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGCGTCTGTGGAAGCGTGGCTGCTATGC 480  
|||||  
Db 141 SerAspGlySerSerLeuThrThrTrpGluThrProLeuGlyAsnValGlyAlaLeuCys 160  
QY 481 TGTGTGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
|||||  
Db 161 CysTrpGluHisLeuGlnProLeuSerArgTyThrAlaMetTyrSerGlnHisGluGluIle 180  
QY 541 CATTTGCGCGCTTGGCCGAGCTTTAGCCCTTATCTAATGCGGCGCAAGCCCTGGGCCCT 600  
|||||  
Db 181 HisIleAlaIaIaTrpProSerPheSerLeuTyThrSerAlaThrAlaIaLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCCTTGAAGGCGCAATGCTTCTGTAACG 660  
|||||  
Db 201 AspValAsnThrAlaIaIaSerArgLeuTyThrAlaIaGluGlyGlnCysPheValIleAla 220  
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTATGCACAAATGAAGAGCAT 720  
|||||  
Db 221 ProCysAlaValValSerAspGluMetIleAspLeuLeuCysProAspAspArgArg 240  
QY 721 GCGTTGCTTCTGCTGTGGTGGAGACACTACGATCATAGGGCTGATGGTGTGACTTG 780  
|||||  
Db 241 AlaLeuLeuSerAlaGlyGlyGlyHisAlaArgIleTyrglyProAspGlyArgGluLeu 260  
QY 781 GTCGGCGCTTGTCCGAAATGAAGAGGATATCTCTACGCAACCTGATCCTCGAGTA 840  
|||||



Db 261 ValThrProleuGlyGluAsnGluGlyLeuLeuIleAlaGluLeuAspSerAlaIa 280  
QY 841 CGCATCCTTGCTAAATGCGGAGACCCGTGCTCATTTATCCGTCGACATTACT 900  
Db 281 IleThrPheAlaIaLeuAlaIaAspProValGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAATTACCGTAGTTGAA 942  
Db 301 ArgLeuLeuPheAsnProSerAlaAsnLysThrValIleLys 314  
RESULT 8  
Q6RWS5\_9ZZZZ PRELIMINARY; PRT; 345 AA.  
AC Q6RWS5;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE NitriIase (EC 3.5.5.7).  
GN ORFNames=BD7611;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxId=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriIase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487427; AAR97374.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitriIase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 345 AA; 37242 MW; 5F8A3CFE530D2F1D CRC64;  
Alignment Scores:  
Pred. No.: 6.92e-88 Length: 345  
Score: 1081.00 Matches: 201  
Percent Similarity: 78.6% Conservative: 41  
Best Local Similarity: 65.3% Mismatches: 66  
Query Match: 59.2% Indels: 0  
DB: 2 Gaps: 0  
US-09-751-299-3 (1-1014) x Q6RWS5\_9ZZZZ (1-345)  
QY 16 AAGTCGCTGCGTGAAGCCGCCCATCTACATGATTTGGAGCGGACGATGACAA 75  
Db 9 ArgValAlaAlaValAlaGlnAlaIaProGluPheLeuAsnLeuGluAlaThrValAspLys 28  
QY 76 ACCATTGAGTTGATGGAAGAGACGACGTAATATGCTGCTGTGATCGCCTTCCGGA 135  
Db 29 ThrIleAlaLeuIleGluGlnAlaIaIaArgGlyGlyAlaSerLeuIleAlaPheProGlu 48  
QY 136 ACTTGATTCCAGGCTACCCAGTGTCTTTGGCTTGACTCACGACGATGGGCAATGCA 195  
Db 49 ThrTyrIleProGlyTyrProTyrPheAlaIaIaProIleTyrGlyMetLys 68  
QY 196 TTGTAGCCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCTCAAGCTAAGCGATT 255  
Db 69 PheIleGlnAlaTyrHisAspAsnSerMetValIleAspGlyAlaGlnPheGluArgIle 88  
QY 256 TCAGATGACGCAAGCGGTTGGGAATCATGTGTCACCGCTGGGATGAGTGAACGGGTGGT 315  
Db 89 AlaGlnAlaIaIaSerArgCysAsnIleThrValValLeuGlyPheSerGluLysAspAla 108

QY 316 GGCACCTTTACATCAGTCAGTGGTTTCATAGCCGATAATGTGACACCATTTGGGCCCCG 375  
Db 109 GlySerLeuTyrIleAlaGlnAlaIleLeuSerProGluGlyLysThrIleAlaThrArg 128  
QY 376 CGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTCGGCGAAGGGATGTTTCATCG 435  
Db 129 ArgLysLeuLysProThrHisValGluArgAlaIlePheGlyGluGlyAspGlySerAsp 148  
QY 436 CTAGCGTTTTCAGACGCTGTGTTGGAAGGCTGGGTGCTTATGCTGTGGAGACACTT 495  
Db 149 LeuAlaValHisAspThrLysLeuGluArgValGlyAlaLeuCysCysTrpGluHisLeu 168  
QY 496 CAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGATTCATTGTGCGGCTTGG 555  
Db 169 GlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisIleAlaIaIaTrp 188  
QY 556 CCGAGCTTTAGCCTTATCTTATGCGCGGAAAGCCCTGGGGCCTGATGTCATGTAGCG 615  
Db 189 ProSerPheSerLeuTyrValAspAlaIaIaTyrAlaLeuGlyProGluValAsnAsnAla 208  
QY 616 GCCTCTCGAATCTATGCCGTTGAAGGCAATGCTTGTACTAGCGTGTGCGCTGCT 675  
Db 209 AlaSerArgLeuTyrAlaValGluGluGlnCysPheValValAlaProCysAlaThrVal 228  
QY 676 TCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTGCT 735  
Db 229 SerGlnLysMetIleAspMetLeuCysGluThrProGluGlnAlaLeuLeuLysPro 248  
QY 736 GGTGTGACACTCACGATATAGGCGCTGATGCTGTGACTTGTGCGCCTTGGCC 795  
Db 249 GlyGlyGlyHisAlaGlnIleTyrGlyProAspGlyArgSerLeuAlaAspProLeuPro 268  
QY 796 GAAATGAAGAGGTATTTCTTACGCAACCTTGATCTTGAGTACGCATCTTGCTTAA 855  
Db 269 ProAspAlaGluGlyLeuLeuTyrAlaAspIleAspLeuAlaIaIleThrLeuAlaLys 288  
QY 856 ATGGCGGAGACCCGTGCTGCTATTTCCCGTCCGACGATTAATCTGCTGCTAATAGAT 915  
Db 289 AlaAlaIaIaAspProAlaGlyHisTyrSerArgProAspValThrGlnLeuLeuAsp 308  
QY 916 GCGAGCCCTAATTAACCGGTAGTT 939  
Db 309 ArgAsnProLysProArgValVal 316  
RESULT 9  
Q6RWE3\_9ZZZZ PRELIMINARY; PRT; 334 AA.  
ID Q6RWE3\_9ZZZZ  
AC Q6RWE3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE NitriIase (EC 3.5.5.7).  
GN ORFNames=BD7051;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxId=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriIase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487559; AAR97506.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitriIase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlase/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.





Db 89 LysLeuArgAlaAlaAlaLysArgAsnLysMetValValLeuGlyLeuSerGluArg 108  
QY 310 GTGGTGGCAACCTTTACATCAGTGGTTCATAGCGGATAAGTGACACCATTTGG 369  
Db 109 AspGlyGlySerLeuTyrIleAlaGlnTPrIleIleGlyProAspGlyGluThrIleAla 128  
QY 370 GCCCGGCGAAAGTTGAAACCTACTTTTGTGAACGTAATTGTTGCGGGAAGGGATGGT 429  
Db 129 LysArgArgLysLeuLysProThrIleAlaGluArgThrValPheGlyGlyAspGly 148  
QY 430 TCATCGCTAGCCGTTTTCGAGACGTCGTGTGAAGGCTGGGCTTATGCTGTGGAG 489  
Db 149 SerHisLeuAlaValHisGluLeuAspValGlyArgLeuGlyAlaLeuCysCysTPrGlu 168  
QY 490 CACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATTCAATTGTCG 549  
Db 169 HisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisValAla 188  
QY 550 GCTTGGCCGAGCTTTAGCCCTTATTCCTAATGCGCGGAAAGCCCTGGGCGCTGATTCAT 609  
Db 189 AlaTPrProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGluValAsn 208  
QY 610 GTAGCGGCTCTCGAATCTATGCGCGTGAAGGCAATGCTTGTACTAGCGCTGTGCG 669  
Db 209 AsnAlaAlaSerLysIleTyrAlaValGluGlySerCysPheValIleAlaProCysAla 228  
QY 670 CTGCTTCAATCCATGATCGATGCTTGTGTACAGATGACGAAAGCATGCGCTTGT 729  
Db 229 ThrValSerGlnAlaMetIleAspGluLeuCysAspThrProGluLysHisGlnPheLeu 248  
QY 730 CTGGCTGGTGGTGACACTCAACGATCATAGGGCGCTGATGGTGACTTGGTCGGCCT 789  
Db 249 HisAlaGlyGlyGlyPheAlaValIleTyrGlyProAspGlyAlaProLeuAlaAlaPro 268  
QY 790 CTTCGCCGAAATGAAGAGGGTATTCTTACGCAAACTTGATCCTGAGTACGCACTTCT 849  
Db 269 LeuProProAspLysGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIleSerVal 288  
QY 850 GCTAAATGCGCGCAGACCCCTGCTGTCTATTATCCCGCCGACATTACTCGCTTGCTA 909  
Db 289 AlaLysAlaAlaAlaAspProAlaGlyHisIleTyrAlaArgProAspValThrArgLeuLeu 308  
QY 910 ATAGATCGCAGCCCT 924  
Db 309 PheAsnAsnArgPro 313  
RESULT 11  
Q6RWS2\_9ZZZZ  
ID Q6RWS2\_9ZZZZ PRELIMINARY; PRT; 348 AA.  
AC Q6RWS2;  
DT 05-JUL-2004 (TReMBLrel. 27, Created)  
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5338;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487430; AAR97377.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydase.

DR InterPro; IPR002332; PII\_GlnB\_UMP\_S.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN 1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN\_1.  
KW Hydrolase.  
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;  
Alignment Scores:  
Pred. No.: 4.05e-84 Length: 348  
Score: 1039.00 Matches: 191  
Percent Similarity: 78.2% Conservative: 46  
Best Local Similarity: 63.0% Mismatches: 66  
Query Match: 56.9% Indels: 0  
DB: 2 Gaps: 0  
US-09-751-299-3 (1-1014) x Q6RWS2\_9ZZZZ (1-348)  
QY 16 AAGTCGCGTCGTCGCAAGCCGCCGATCTACATGATTTGGAGGCGACGGTGACAA 75  
Db 8 ArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeuGlyAlaIleSerLys 27  
QY 76 ACCATTGAGTTGATGAGAAGACGACGATTAATGCTGCTGATCGCTTCCGAA 135  
Db 28 GlyLeuSerLeuIleGluGluAlaAlaSerAsnGlyAlaLysLeuIleAlaPheProGlu 47  
QY 136 ACTTGATTCAGGCTACGCCATGCTGTTCTTGGCTGACTCACGACATGGGCAATGCAA 195  
Db 48 ThrTPrIleProGlyTyrProTPrIleTPrIleTPrLeuAspSerProAlaTPrGlyMetArg 67  
QY 196 TTTGTACGCCAATACGATGAGAATCTATGTGAGTTGATGGCCCTCAAGCTAAGCGCAT 255  
Db 68 PheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySerGluGlnAlaLysArgMet 87  
QY 256 TCAGATGCAGCCCAAGCGGTTGGGAATCATGTGTCACCCCTGGGATGATGAACGGGTCGT 315  
Db 88 AsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGlyTyrSerGluArgSerGly 107  
QY 316 GGCACCCCTTACATGATGATGCTGTTCATAGGCGATAATGTTGACACCATTTGGGCCCG 375  
Db 108 GlySerLeuTyrMetGlyGlnSerIleIleAsnAspLysGlyGluThrIlePheThrArg 127  
QY 376 CGAAAGTTGAACCTACTTTTGTGAACGTAATTGTTGCGGGAAGGAGATGTTTCATCG 435  
Db 128 ArgLysLeuLysProThrHisValGluArgThrValPheGlyGlyGlyAspGlySerHis 147  
QY 436 CTAGCGGTTTTCGAGACGCTGTGTGAAGGCTGGGCTTATGCTGTGGAGACACTT 495  
Db 148 LeuCysValMetAspThrGluIleGlyArgValGlyAlaMetCysCysTPrGluHisLeu 167  
QY 496 CAACCGCTAACAAATACGCTTGTGATGCACAAATGAAGAGATTCAATTGCGGCTTGG 555  
Db 168 GlnProLeuSerLysTyrAlaMetTyrSerGlnAspGluGlnIleHisIleAlaSerTPr 187  
QY 556 CCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCGCTGATGCAATGTAGCG 615  
Db 188 ProSerPheSerLeuTyrArgGlyAlaIleTyrAlaLeuGlyProGluLeuAsnAlaAla 207  
QY 616 GCCTCTCGAATCTATGCCGTTGAAGGCGCAATGCTTGTACTAGCGTGTGCGCTCGTT 675  
Db 208 AlaSerGlnMetTyrAlaAlaGluGlyGlnCysPheValLeuAlaProCysAlaThrVal 227  
QY 676 TCACAATCCATGATCGATATGCTTGTGTAGAGATGACGAAAGCATGCGTGTCTGGCT 735  
Db 228 SerLysGluMetIleGluMetLeuIleAspAspProArgLysGluProLeuLeuGlu 247  
QY 736 GGTGTGACACTCACGATATCATTAGGCGCTGATGCTGTGACTTGTGCGGCTTGGCC 795  
Db 248 GlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgProLeuAlaLysProLeuPro 267  
QY 796 GAAATGAAGAGGTAATTCTTACGCAAACTTGATCTGAGTAGCGATCTTGCTTAA 855

Db 268 GluAsnGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIleSerMetAlaLys 287  
QY 856 ATGGCGGACAGCCCTGCTGTCATTATTCGCCGCCGACATTACTCGCTTGCTAATAGAT 915  
Db 288 AlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspValThrArgLeuPheAsn 307  
QY 916 CGCAGCCCT 924  
Db 308 SerAlaPro 310

RESULT 12

Q6RWF6\_9ZZZZ PRELIMINARY; PRT; 336 AA.  
ID Q6RWF6\_9ZZZZ

AC Q6RWF6; 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5295;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RA "Exploring nitrilase sequence space for enantioselective catalysis.";  
RT Appl. Environ. Microbiol. 70:2429-2436(2004).  
RL EMBL; AY487546; AAR97493.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS00263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
KM Hydrolase.  
SQ SEQUENCE 336 AA; 36221 MW; CC370C188F94CB72 CRC64;

Alignment Scores:  
Pred. No.: 6.08e-84 Length: 336  
Score: 1037.00 Matches: 197  
Percent Similarity: 74.1% Conservative: 46  
Best Local Similarity: 60.1% Mismatches: 73  
Query Match: 56.8% Indels: 12  
DB: 2 Gaps: 3

US-09-751-299-3 (1-1014) x Q6RWF6\_9ZZZZ (1-336)

QY 16 AAGTCGCTGCGTGAAGCCGCCGATCTACATGATTGGAGCGACGCTGACAAA 75  
Db 9 LysAlaAlaValGlnAlaGlyProAlaPheLeuAspGlyGlyValGluArg 28  
QY 76 ACCATTGAGTTGATGAAGACGACGATATATGCTGCTGATGCGCTTCCGGA 135  
Db 29 AlaValSerLeuIleGlyGlnAlaAlaAlaGluGlyAlaGlnLeuIleAlaPheProGlu 48  
QY 136 ACTTGATTCCAGGCTACCACTGTTCTTTGGCTTGACTGACGACATGGCAATGCAA 195  
Db 49 ThrTrpIleProGlyTyrProTrpHisThrTrpLeuGlySerProAlaTrpAlaMetGlu 68  
QY 196 -----TTGTACGCCAATACATGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAG 249  
Db 69 LysGlyPheValGlnArgTyrPheAspAsnAlaLeuArgHisGlySerProGlnAlaGlu 88  
QY 250 CGCATTTTCAGATGCAGCCCAAGCGGTTGGGAATCATGCTACCCCTGGGATGAGTGAACGG 309  
Db 89 ArgIleSerGlyAlaAlaAlaGluHisLysIleMetValSerLeuGlyPheAlaGluArg 108

QY 310 GTCGGTGACACCTTACATCAGTCAGTGTTTCATAGGCGAATAAGTGACACCATGGG 369  
Db 109 AspGlyGlyThrLeuTyrIleAlaGlnTrpLeuIleGlyProAspGlyGlnThrIleSer 128  
QY 370 GCCCGCGGAAGTTGAACCTACTTTGTTGAACGCTACTTTGTTGGCGAAGGATGGT 429  
Db 129 ArgArgArgLysLeuLysProThrHisValGluArgThrValPheGlyGluGlyAspGly 148  
QY 430 TCATCGCTACGGTTTTTCGAGACGTCCTGTGGAAGGCTGGGTGCTTATGCTGTGGAG 489  
Db 149 SerAspLeuSerValHisAspThrAlaLeuGlyArgIleGlySerLeuCysCysTrpGlu 168  
QY 490 CACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGATTCATTGTGCG 549  
Db 169 HisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnIleHisIleGly 188  
QY 550 GCTTGCGCGAGCTTTAGCCTTTATCCTTAATGCCGCGAAGCCCTGGGCTGATGCAAT 609  
Db 189 AlaTrpProSerPheSerLeuTyrGlnProPheAlaAsnAlaLeuSerProGluValAsn 208  
QY 610 GTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTAACGCTGTCGCG 669  
Db 209 IleAlaValSerArgValTyrAlaValGluGlyGlnCysPhePheLeuAlaProCysAla 228  
QY 670 CTCGTTTCAATTCATGATCGATATGCTTTGTAACAGATGACGAAAGCATGCGTTCCT 729  
Db 229 ThrValSerAspAlaMetIleGluThrLeuCysAspThrProGluLysGlnGlyLeuIle 248  
QY 730 CTGCTGTGTGTGACACTCAGCTATGATGAGCGCTGATGCTGCTGCTGCGGCT 789  
Db 249 ArgAlaGlyGlyGlyHisAlaAlaIlePheGlyProAspGlySerLeuLeuThrProThr 268  
QY 790 CTTGCCGAATGAAGGGTATTTCTTCGCAAACTTGATCCTGAGTACGCATCCTT 849  
Db 269 ValAlaAspThrTyrGluGlyLeuLeuTyrAlaGluLeuAspLeuGlyValIleSerIle 288  
QY 850 GCTAAATGCGCGGACAGCCCTGCTGCTGCTATTCCTCCGACATTAATCGCTTGCTA 909  
Db 289 AlaLysSerAlaAlaAspProAlaGlyHisTyrSerArgProAspValThrArgLeuLeu 308  
QY 910 ATAGATCGACCCCTAAA-----TTACCG-----GTAGTT 939  
Db 309 LeuAsnGlnThrProSerLysArgValGlnAsnMetValLeuProLeuGluThrValThr 328  
QY 940 GAAATTGAAGTGATCTTCGTCCT 963  
Db 329 GluProGluGlyProValGlnPro 336

RESULT 13

Q6RWF0\_9ZZZZ PRELIMINARY; PRT; 353 AA.  
ID Q6RWF0\_9ZZZZ

AC Q6RWF0; 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RA "Exploring nitrilase sequence space for enantioselective catalysis.";  
RT Appl. Environ. Microbiol. 70:2429-2436(2004).  
RL EMBL; AY487552; AAR97499.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.



DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; Ntlase/CNhydctse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 353 AA; 38321 MW; 2AA05257B3A7BC6C CRC64;

Alignment Scores:

Pred. No.:	6.14e-84	Length:	353
Score:	1037.00	Matches:	195
Percent Similarity:	74.7%	Conservative:	50
Best Local Similarity:	59.5%	Mismatches:	71
Query Match:	56.8%	Indels:	12
DB:	2	Gaps:	3

US-09-751-299-3 (1-1014) x Q6RWF0\_9ZZZZ (1-353)

QY 13 ATCAAGTTCGCGTGGTGCAGACCGCCCGATCTACATGATTTGGAGCGACGGTGAC 72  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 8 LeuLysValAlaIaValGlnAlaIaIaProIaPheLeuAspLeuAspIaSerIleGlu 27  
QY 73 AAAACCATGTGATGATGGAAGAAGCAGCAGCTAATATGCTCTGCTGATCGCTTTCCG 132  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 28 LysThrIleArgTyrValAspGluAlaIaIaIaIaGlyAlaIaLysLeuIleAlaPhePro 47  
QY 133 GAAACTTGATTCAGGCTACCCATGTTTCTTGCTTGACTCACCAGCATGGGCATG 192  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 48 GluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuGlyAlaProIaIaTrpAlaIle 67  
QY 193 CAA-----TTTGTACGCCAATACCATGAGAATCTCATTTGAGTTGGATGGCCCTCAAGCT 246  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 68 MetArgGlyPheValSerArgTyrPheAspAsnSerLeuGlnTyrGlySerProGluAla 87  
QY 247 AAGCGCATTTGAGATGACGCCAAGCGGTTGGGAATCATGTCACCCCTGGGAGTGAAGTGA 306  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 88 GluArgLeuArgAspAlaIaIaArgArgAsnLysIleTyrIleAlaLeuGlyLeuSerGlu 107  
QY 307 CGGTCGGTGACCCCTTTACATCATGTCAGTGTTCATAGCCGATATGTTGACACCATT 366  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 108 ArgAspGlyGlySerLeuTyrIleAlaGlnTrpIleIleGlyProGlyGlyIuThrVal 127  
QY 367 GGGGCGGCGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTTGGCGAAGGGGAT 426  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 128 AlaGlnArgArgLysLeuLysProThrHisAlaGluArgThrValPheGlyGluGlyAsp 147  
QY 427 GGTTCATCGCTAGCGGTTTTCGAGACGCTGTTGGAAGGCTGGGCTTATGCTGTGG 486  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 148 GlySerHisLeuAlaValHisAspLeuAspIleGlyArgLeuGlyAlaLeuCysCysTrp 167  
QY 487 GAGCACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATAGAAAGATTCATGTT 546  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 168 GluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnIleHisVal 187  
QY 547 GCGGCTTGGCCGAGCTTTAGCCTTTATCTTAATGCCGCGAAAGCCCTGGGGCTGATGTC 606  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 188 AlaAlaIaTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGluVal 207  
QY 607 AATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGCAATGCTTCGTAAGCGTGTGT 666  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 208 AsnAsnAlaIaSerLysIleTyrAlaValGluGlySerCysPheValIleAlaProCys 227  
QY 667 GCGCTGTTTCACAATCCATGATGATGCTTTGTATACAGATGACGAAGAAGCATGCGTTG 726  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 228 AlaThrValSerGlnValMetIleAspGluLeuCysAspThrProGluLysHisGlnPhe 247  
QY 727 CTTCTGGCTGGTGGACACTCACGTAATAGAGGCTGATGGTGGTGAAGCTTGTGTCGG 786  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 248 LeuHisValGlyGlyPheAlaValIleTyrGlyProAspGlySerProLeuAlaLys 267  
QY 787 CTTCTTGCAGAAATGAAGAGGTATTCTTACGCAACCTTGATCTGAGTACGATC 846  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 268 ProLeuProProAspGlnGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIleSer 287  
QY 847 CTTGCTAAATGCGCGCAGACCCCTGCTGCTATTATTCGCCGTCGACATTACTGCTTG 906  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 288 ValAlaLysAlaIaIaAspProAlaGlyHisTyrAlaArgProAspValThrArgLeu 307  
QY 907 CTAATAGATCGACCCCT-----AAATTACCGGTA----- 936  
|||:::|||||  
Db 308 LeuPheAsnAsnArgProAlaAsnArgValGluLysLeuAlaLeuProValAspGlnGlu 327  
QY 937 GTTGAATTGAAGCTGATCTTCGT 960  
Db 328 AlaGluValAspSerProLeuLys 335  
RESULT 14  
Q6RWS3\_9ZZZZ  
ID Q6RWS3\_9ZZZZ PRELIMINARY; PRT; 337 AA.  
AC Q6RWS3;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7048;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487429; AAR97376.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydratase.  
DR InterPro; IPR003010; Ntlase/CNhydctse.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 337 AA; 36423 MW; 841BABA391031A81 CRC64;  
Alignment Scores:  
Pred. No.: 1.39e-83 Length: 337  
Score: 1033.00 Matches: 192  
Percent Similarity: 76.2% Conservative: 55  
Best Local Similarity: 59.3% Mismatches: 67  
Query Match: 56.6% Indels: 10  
DB: 2 Gaps: 3  
US-09-751-299-3 (1-1014) x Q6RWS3\_9ZZZZ (1-337)  
QY 16 AAGTTCGCTGCGTGCAGACCGCCCGATCTACATGATTTGGAGCGGACGGTGACAA 75  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 9 LysValAlaValValGlnAlaIaIaProIaIaTrpLeuAspLeuAspIaSerIleAlaLys 28  
QY 76 ACCATTGAGTTGATGGAAGAAGCAGCAGTAATATGCTGCTGATGCGCTTCCGAA 135  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
Db 29 SerIleAlaLeuIleGluGluAlaIaIaAsnGlyAlaLysLeuIleAlaPheProGlu 48  
QY 136 ACTTGATTCAGGCTACCAATGCTTTCTTTGGCTTGACTCACGACATGGGCAATG--- 192  
Db 49 AlaPheIleProGlyTyrProTrpTyrIleTrpLeuAspSerProAlaIaIleGly 68  
QY 193 ---CAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAG 249  
Db 69 ArgGlyPheValGlnArgTyrPheAspAsnSerLeuAlaTyrAspSerProGlnAlaGlu 88



[illegible][illegible]

Db	249	LeuHisValGlyGlyHisAlaIlePheGlyProAspGlySerAlaIleAlaAsp	268
QY	787	CCTCTTGCCGAAATGAAGGGTATTTCTACGCAACCTTGATCCTGAGTACGCATC	846
Db	269	LysLeuProSerAspGlnGlyLeuLeuPheAlaAspIleAspLeuGlyAlaIleGly	288
QY	847	CTTGCTAAATGGCGGACAGCCCTGCTGGTCATTATCCCGTCCCGACATTACTGCGCTTG	906
Db	289	IleAlaLysAsnAlaAlaAspProAlaGlyHisTyrSerArgProAspValThrArgLeu	308
QY	907	CTAATGATCGCAGCCCTAA-----TTACCGGTAGTTGAAATT	945
Db	309	LeuLeuAsnLysLysProSerLysArgValGluHisPheAlaLeuProLeuAspThrLeu	328
QY	946	GAAGGTGAT	954
Db	329	AlaGlyGlu	331

Search completed: April 27, 2006, 01:35:26  
Job time : 256.429 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:06:26 ; Search time 33.8987 Seconds  
(without alignments)  
2628.598 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1826  
Sequence: 1 atgaagaagactatcaagt.....cgcaactcgagaatttga 1014

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 segs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB/spool/US09751299/runat\_26042006\_090032\_18619/app\_query.fasta\_1  
-DB=A\_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs02p  
-USER=US09751299 @CGN\_1\_1\_476 @runat\_26042006\_090032\_18619 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A\_Geneseq\_21: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*  
9: geneseqp2005s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765	96.7	337	4	AAE05489	Aae05489 Nitriase
2	1761	96.4	337	7	ADC24119	Adc24119 Protein s
3	1761	96.4	337	8	ADH36220	Adh36220 Chemical
4	1761	96.4	337	8	ADG93919	Adg93919 Nitriase
5	1761	96.4	337	8	ADI62518	Adi62518 Nitriase
6	1761	96.4	337	8	ADI64639	Adi64639 Nitriase
7	1339	73.3	337	7	ADC23791	Adc23791 Protein s
8	1339	73.3	337	7	ADC23781	Adc23781 Protein s
9	1339	73.3	337	8	ADH35882	Adh35882 Chemical

10	1339	73.3	337	8	ADH35892	Adh35892 Chemical
11	1339	73.3	337	8	ADG93583	Adg93583 Nitriase
12	1339	73.3	337	8	ADG93593	Adg93593 Nitriase
13	1339	73.3	337	8	ADI62180	Adi62180 Nitriase
14	1339	73.3	337	8	ADI62190	Adi62190 Nitriase
15	1339	73.3	337	8	ADI64311	Adi64311 Nitriase
16	1339	73.3	337	8	ADI64301	Adi64301 Nitriase
17	1139	62.4	333	7	ADC24025	Adc24025 Protein s
18	1139	62.4	333	8	ADH36126	Adh36126 Chemical
19	1139	62.4	333	8	ADG93827	Adg93827 Nitriase
20	1139	62.4	333	8	ADI62424	Adi62424 Nitriase
21	1139	62.4	333	8	ADI64545	Adi64545 Nitriase
22	1136	62.2	332	7	ADC23873	Adc23873 Protein s
23	1136	62.2	332	8	ADH35974	Adh35974 Chemical
24	1136	62.2	332	8	ADG93675	Adg93675 Nitriase
25	1136	62.2	332	8	ADI62272	Adi62272 Nitriase
26	1136	62.2	332	7	ADG93827	Adg93827 Nitriase
27	1091	59.7	355	7	ADC24039	Adc24039 Protein s
28	1091	59.7	355	8	ADH36140	Adh36140 Chemical
29	1091	59.7	355	8	ADG93841	Adg93841 Nitriase
30	1091	59.7	355	8	ADI62438	Adi62438 Nitriase
31	1091	59.7	355	8	ADI64559	Adi64559 Nitriase
32	1081	59.2	345	7	ADC24067	Adc24067 Protein s
33	1081	59.2	345	8	ADH36168	Adh36168 Chemical
34	1081	59.2	345	8	ADG93868	Adg93868 Nitriase
35	1081	59.2	345	8	ADI62466	Adi62466 Nitriase
36	1081	59.2	345	8	ADI64587	Adi64587 Nitriase
37	1051	57.6	334	7	ADC23739	Adc23739 Protein s
38	1051	57.6	334	8	ADH35840	Adh35840 Chemical
39	1051	57.6	334	8	ADG93541	Adg93541 Nitriase
40	1051	57.6	334	8	ADI62138	Adi62138 Nitriase
41	1051	57.6	334	8	ADI64259	Adi64259 Nitriase
42	1042	57.1	353	7	ADC23945	Adc23945 Protein s
43	1042	57.1	353	8	ADH36046	Adh36046 Chemical
44	1042	57.1	353	8	ADG93747	Adg93747 Nitriase
45	1042	57.1	353	8	ADI62344	Adi62344 Nitriase

ALIGNMENTS

RESULT 1	
AAE05489	
ID AAE05489	standard; protein; 337 AA.
XX	
AC AAE05489;	
XX	
DT 25-SEP-2001	(first entry)
XX	
DE Nitriase protein #2.	
XX	
KW Nitriase; enantiomer; alpha-substituted carboxylic acid;	
KW stereoselective; hydrolysis; amino nitriole; cyanohydrin.	
XX	
OS Unidentified.	
XX	
PN W0200148175-A2.	
PN	
PD 05-JUL-2001.	
XX	
PF 29-DEC-2000; 2000WO-US035555.	
XX	
PR 29-DEC-1999; 99US-0173609P.	
PR 07-DEC-2000; 2000US-0254414P.	
XX	
PA (DIVE-) DIVERSA CORP.	
PA (MADD/) MADDEN D.	
XX	
PI Madden M, Weiner DP, Chaplin JA;	
XX	
DR WPI; 2001-465211/50.	
DR N-PSDB; AAD11167.	
XX	
PT Producing alpha-substituted carboxylic acid enantiomers by contacting	



PT aldehyde or ketone with cyanide, ammonia compound or its salt or amine  
PT and hydrolyzing the resulting amino nitrile or cyanohydrin with  
PT Nitrilase.  
XX  
PS Claim 30; Page 86-87; 87pp; English.  
XX  
CC The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolysing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a nitrilase protein #2  
XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 6.87e-185 Length: 337  
Score: 1765.00 Matches: 337  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.7% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x AAE05489 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 MettysgluAlaIeIyValAlaCysValGlnAlaIaProIleTyrMetAspLeuGlu 20  
QY 61 GCGACGGTGGACAAACCATTTGATTGATGAAGAGCAGCAGTAAATATGCTGCTG 120  
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40  
QY 121 ATCGCCTTCCGAAACTTGATTCCAGGCTACCCAGTGTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCATGCAATTGTGACGCCAATACCATGAGAACTCATTTGGAGTTGGAGCCCT 240  
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAACTAAGCGCATTTCAATGACAGCCAGCGGTTGGGAATCATGTGACCCCTGGGGATG 300  
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTCGGTGGCACTTTACATCAGTCAGTGGTTCATAGGGCATATAGGTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyIle 140  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGCTGTTGGAAAGGCTGGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyLeuCys 160  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluIle 180  
QY 541 CATTGTGGGCTTGGCCGAGCTTTAGCCTTTATCCTAATAGCGCGAAAGCCCTGGGCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTAAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGGCGCTGTTTCAACAATCCATGATCGATATGCTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240

QY 721 GCGTGTCTTCTGGCTGTGTGTGACACCTCAGTATCATAGGGCCGTGATGCTGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspArgIleGlyAspLeu 260  
QY 781 GTCCGCGCTTGGCCGAAATGAAGAGGGTATTCTTACGCAACCTGATCCTGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGTGTAATAGCGCGAGACCCCTGCTGTCATTATTCCTCCGACATTACT 900  
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAAATTGAAGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 2  
ADC24119  
ID ADC24119 standard; protein; 337 AA.  
XX  
AC ADC24119;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Protein sequence (SeqID 386) exhibiting nitrilase activity.  
XX  
KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO2003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
XX  
DR WPI; 2003-201417/19.  
DR N-PSDB; ADC24118.  
XX  
XX

CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and

CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.

XX Sequence 337 AA;

SO Alignment Scores:

Pred. No.:	1.9e-184	Length:	337
Score:	1761.00	Matches:	336
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.7%	Mismatches:	0
Query Match:	96.4%	Indels:	0
DB:	7	Gaps:	0

US-09-751-299-3 (1-1014) x ADC24119 (1-337)

```
OY 1 ATGAAGAAGCTATCAAGTCGCCCTGCGTGCAAGCCGCCCGATCTACATGATTGGAG 60
    |||||||
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaAlaProIleTyrMetAspLeuLys 20
OY 61 GCGACGGTGGACAAACCATTTGAGTTGATGAGAGACGACGTAATATGCTGCTGTG 120
    |||||||
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40
OY 121 ATGCGCTTTCGGAACTTGAGTTCCAGGCTACCCATGCTTCTTTGGCTTGACTCACA 180
    |||||||
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
OY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGAGGCCCT 240
    |||||||
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80
OY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAAAGCGGTTGGGAATCATGGTCACCCCTGGGATG 300
    |||||||
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100
OY 301 AGTGAACGGGTCCGTGGGCACCCTTTACATCAGTCAGTGTTTATAGGCGATAATGGTGAC 360
    |||||||
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120
OY 361 ACCATTGGGGCCCGGCGAAAGTTGAACCTACTTTTGTGAACTGACTTTGTTGGCGGAA 420
    |||||||
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140
OY 421 GGGGATGTTTCATCGCTGAGCGGTTTTTCGAGACGTCCTGTTGAAAGGCTGGTGCTTATGC 480
    |||||||
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160
OY 481 TGTGGGAGCACCTTCAACCGCTTAACAATAACGCTTTGTATGCACAAATGAGAAGATT 540
    |||||||
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180
OY 541 CATTTGCGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCCT 600
    |||||||
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200
OY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGGCAATGCTTCTGACTAGCG 660
    |||||||
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220
OY 661 TCGTGTGGCTCGTTTCAACATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720
    |||||||
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240
OY 721 GCGTTGCTTCTGGCTGGTGGTGGACACTCACGTAATCATAGGGCCTGATGGTGGACTTG 780
    |||||||
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260
OY 781 GTCGCGCCTCTTGGCCGAAATGAAGAGGGTATTCTCTACGCCAAACCTTGATCTCGAGTA 840
    |||||||
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280
OY 841 CGCATCCTTGCTAAATGGCGGCGAGACCCGTGCTGTCATTATTCCCGTCCCGACATTACT 900
    |||||||
```

```
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300
OY 901 CGCTTGCTAATAGATCGAGCCCTTAATTACCGGTAGTTGAATTTGAAGGTGATCTTGT 960
    |||||||
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGlyAspLeuArg 320
OY 961 CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGGCGAACTCGAAGAATTT 1011
    |||||||
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337
```

RESULT 3

ADH36220  
ID ADH36220 standard; protein; 337 AA.

XX ADH36220;

DT 11-MAR-2004 (first entry)

DE Chemical process monitoring-related nitrilase protein sequence SeqID386.

KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.

XX Unidentified.

XX WO2003098187-A2.

XX 27-NOV-2003.

XX 15-MAY-2003; 2003WO-US015639.

PR 15-MAY-2002; 2002US-0380737P.

PA (DIVE-) DIVERSA CORP.

XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;

PI Mcquaid J, Stege J;

XX WPI; 2004-142708/14.

DR N-PSDB; ADH36219.

PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.

PS Claim 73; SEQ ID NO 386; 277pp; English.

XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.

XX Sequence 337 AA;

SO Alignment Scores:

Pred. No.:	1.9e-184	Length:	337
Score:	1761.00	Matches:	336
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.7%	Mismatches:	0
Query Match:	96.4%	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-3 (1-1014) x ADH36220 (1-337)

```
OY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTGCAAGCCGCCGATCTACATGATTGGAG 60
    |||||||
```

Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaAlaProIleTyrMetAspLeuLys 20

QY 61 GCGACGGTGGACAAACCATTTGAGTTGATGAGAAAGACAGACGTAATATGCTGCTGTG 120

Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40

QY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCCATGTGTTCTTTGGCTTGACTCACCA 180

Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60

QY 181 GCATGGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTGAGATTGGATGGCCCT 240

Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80

QY 241 CAAGCTAAGCCGATTTTCAGATGCAGCCCAAGCCGTTGGGAATCATGTGTCACCCCTGGGATG 300

Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100

QY 301 AGTGAACGGGTGGTGACCCCTTTACATCAGTGTGTTGTAACGCTACTTTGTTGGCGGAA 360

Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120

QY 361 ACCATTGGGGCCCGCGGAAAGTTGAACCTTACTTTGTGTAACGCTACTTTGTTGGCGGAA 420

Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140

QY 421 GGGATGGTTCATCGCTAGCGGTTTTCGAGACGCTGTGGAAGCGTGGGCTTATGTC 480

Db 141 GlyAspGlySerSerLeuAlaValaPheGluThrSerValGlyArgLeuGlyLeuCys 160

QY 481 TGTGGAGACACCTTCAACCGCTAACAAATAGCCTTGTATGCACAAATAGAGATT 540

Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180

QY 541 CATGTGCGGCTTGCGGAGCTTTAGCCTTATCCTAATGCGCGGAAAGCCCTGGGGCCT 600

Db 181 HisCysAlaAlaIleTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200

QY 601 GATGTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTAAGCG 660

Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValaGluGlyGlnCysPheValLeuAla 220

QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720

Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240

QY 721 GCGTGTCTTGGCTGGTGGTGGACACTCAGCATCATAGGGCCTGATGGTGTGACTTG 780

Db 241 AlaLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyLysAspLeu 260

QY 781 GTGCGGCTCTTGGCCGAAATGAGAGGGTATTTCTCTACGCAAACTTGATCCTGAGTA 840

Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280

QY 841 CGCATCCTTGCTAAATGCGCGGACAGCCCTGCTGTCATTAATCCCGTCCGACATTAAT 900

Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300

QY 901 CGCTTGCTAATAGATCGAGCCCTTAATTAACCGGTAGTTGAATTTGAAGGTGATCTCGT 960

Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320

QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAATTT 1011

Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 4  
ADG93919  
ID ADG93919 standard; protein; 337 AA.  
AC ADG93919;  
XX  
DT 11-MAR-2004 (first entry)

XX DE Nitrilase enzyme amino acid sequence SeqID386.

XX KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;

XX KW enantioselective transformation; enzyme.

XX OS Unidentified.

XX PN WO2003097810-A2.

XX PD 27-NOV-2003.

XX PF 15-MAY-2003; 2003WO-US015712.

XX PR 15-MAY-2002; 2002US-00146772.

XX PR 09-SEP-2002; 2002US-00241742.

XX PA (DIVE-) DIVERSA CORP.

XX PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

XX DR WPI: 2004-090637/09.

XX DR N-PSDB; ADG93918.

XX PT New isolated or recombinant nucleic acid encoding a polypeptide having

XX PS nitrilase activity, useful for screening enantioselective transformation.

XX Claim 44; SEQ ID NO 386; 295pp; English.

XX CC This invention is related to a novel isolated or recombinant nucleic acid

CC encoding a protein having nitrilase activity. Nitrilase's are capable of

CC converting nitrile's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated

CC nitrilase proteins of the invention have increased activity and stability

CC at increased pH and temperature when compared to those conventionally

CC used. In addition, the nucleic acid of the invention is useful for

CC screening enantioselective transformation. The present sequence is that

CC of a nitrilase enzyme of the invention.

XX SQ Sequence 337 AA;

Alignment Scores:

Pred. No.:	1.9e-184	Length:	337
Score:	1761.00	Matches:	336
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.7%	Mismatches:	0
Query Match:	96.4%	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-3 (1-1014) x ADG93919 (1-337)

QY 1 ATGAAGAAGCTATCAAGTTCGCTGCGTGAAGCCGCCCGATCTACATGATTTGGAG 60

Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaAlaProIleTyrMetAspLeuLys 20

QY 61 GCGACGGTGGACAAACCATTTGAGTTGATGGAAGACAGCAGCAGTAATATGCTGCTGTG 120

Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40

QY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCCATGTGTTCTTTGGCTTGACTCACCA 180

Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60

QY 181 GCATGGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTGAGATTGGATGGCCCT 240

Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80

QY 241 CAAGCTAAGCCGATTTTCAGATGCAGCCCAAGCCGTTGGGAATCATGTGTCACCCCTGGGATG 300

Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100

QY 301 AGTGAACGGGTGGTGACCCCTTTACATCAGTGTGTTGATGAGGATTAATGCTGAC 360



Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGTAACGTACTTTGTCGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGAGTGGTTCATCGCTAGCGGTTTTCGAGCGTCTGTGGAAGCGTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATAACGCTTTGTATGCACAAATAGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATTTGCGCGCTTGCCGAGCTTTAGCCTTTATCCTAATGCGCGGAAAGCCCTGGGCCT 600  
Db 181 HisCysAlaAlaIleTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTACTAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGGCGCTGTTTCAACATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGlyLysHis 240  
QY 721 GCGTTCCTTTGCGCTGCTGTGGTGGACACTCAACGTATCATAGGGCGCTGATGGTGAATTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTCCGCGCTCTTGCCGAAATGAAGGGTATTCTTACGCGAAACCTTGATCCTGGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGCTAAATGGCGGAGACCCCTGCTGTCATTATCCCGTCCGACATTACT 900  
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATGAAGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 5

AD162518

ID AD162518 standard; protein; 337 AA.

XX AC AD162518;

DT 22-APR-2004 (first entry)

DE Nitrilase polypeptide #193.

XX Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipaeamic; enzyme.

OS Unidentified.

PN WO2003106415-A2.

PD 24-DEC-2003.

PF 13-JUN-2003; 2003WO-US018840.

PR 13-JUN-2002; 2002US-0389317P.

PR 28-JUN-2002; 2002US-0392944P.

XX (DIVE-) DIVERSA CORP.  
PA  
XX  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
DR WPI; 2004-090821/09.  
DR N-PSDB; AD162517.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
PS  
PS Claim 41; SEQ ID NO 386; 253pp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrilase polypeptide obtained from an  
CC environmental sample.  
XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 1.9e-184 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 8 Gaps: 0

US-09-751-299-3 (1-1014) x AD162518 (1-337)

QY 1 ATGAAGAAGCTATCAAGGTCGCGCTGCGCAAGCGCCCGCATCTACATGATTTGGAG 60  
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaIleProIleTyrMetAspLeuLys 20  
QY 61 GCGACGGTGGACAAAACCATGAGTTGATGGAAGACGACGCTAATATGCTGCTGTG 120  
Db 21 AlaThrValAspLysThrIleGluLeuMetGluAlaAlaArgAsnAsnAlaArgLeu 40  
QY 121 ATCGCCTTTCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTAGCCCAATACATGAGAACTCATTTGAGTTGATGGCCCT 240  
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGGAATCATGTGTACCCCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGGTGGCACCCCTTACATCAGTCAGTGGTTTCATAGGCGATAATGGTAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGAGTGTTCATCGCTAGCGGTTTTCAGACGCTGCTTGAAGGCTGGGTGCTTATGC 480



Db 141 GLYAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnGluGluIle 180  
QY 541 CATGTGCGGCTTGGCCGAGCTTTAGCCCTTAATCCATAATGCGGCGAAAGCCCTGGGGCT 600  
Db 181 HisCysAlaIaIaIaTrpProSerPheSerLeuTyrProAsnAlaIaIaLeuGlyPro 200  
QY 601 GATGTCATGTCAGCGCCCTCTCGAATCTATGCGCGTTGAAGGCAATGCTTGTACTAGCG 660  
Db 201 AspValAsnValAlaIaIaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240  
QY 721 GCGTTGCTTGTGCTGCTGCTGTGACACTACGATCATAGGGCTGATGGTGACTTGTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlnHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTGCGCGCTTGTGCGGAAATGAAGAGGTTATTTCTACGCAAACTTGATCTTGAAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGTCTAAATGCGGCGAGACCCCTGCTGTCATTATTCCTCCGACATTACT 900  
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlnHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTGGTAAAGCGTCTGAGACGGGTGGCCAACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337  
RESULT 6  
ADI64639  
ID ADI64639 standard; protein; 337 AA.  
XX  
AC ADI64639;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase seq id 194.  
XX  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KW carboxylic acid; cyanohydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KW (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KW & diastereomeric excess; food additive; drug intermediate; nitrilase.  
XX  
OS Unidentified.  
XX  
PN US2004014195-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 15-MAY-2003; 2003US-00440523.  
XX  
PR 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX

PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX  
DR WPI: 2004-121569/12.  
DR N-PSDB; ADI64638.  
XX  
PT Novel isolated or recombinant polypeptide having nitrilase activity,  
PT useful in production of food additives.  
XX  
PS Claim 46; SEQ ID NO 386; 105bp; English.  
XX  
CC The invention describes an isolated or recombinant polypeptide (I)  
CC comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspro.gov/sequence.html?DocID=2004014195, or its  
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining & enantiomeric excess or & diastereomeric  
CC excess. (I) is useful in the production of food additives and drug  
CC intermediates. This is the amino acid sequence of a nitrilase of the  
CC invention.  
XX  
SQ Sequence 337 AA;  
SQ  
Alignment Scores:  
Pred. No.: 1,9e-184 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-3 (1-1014) x ADI64639 (1-337)  
QY 1 ATGAAGAAGCTATCAAGTCGCGCTGCGCAAGCCCGCATCTACATGATTGGAG 60  
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaIaIaProIleTyrMetAspLeuLys 20  
QY 61 GCGACGCTGGAAGAAACCATTTGATGATGAGGAAGACGACGTAATTAATGCTGCTG 120  
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTATGCCCAATACCATGAGAACTCATTTGGATGGCCCT 240  
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTCAAGTCAGCCCAAGCGGTTGGGAATCATGTGCACCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGGCACCCCTTACATCAGTCAGTGTTCATAGGCGATTAATGTTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrLieserGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCGCGCGAAAGTTGAAACCTACTTTGTGAACGTACTTTGTGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGATTCATCGCTAGCGGTTTTCGAGACGTCGTGGAAGGCTGGGCTTATGC 480  
|||||

Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACACTTCAACCGCTAAACAATACGCTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluIle 180  
QY 541 CATTGTGCGCTTGCCGAGCTTTAGCCTTTATCCCTAAATGCGGCGAAGCCCTGGGCGCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGCGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTACTACCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCAAAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240  
QY 721 GCGTTGCTTCTGCGTGTGTGGACACTCACGTAATCATAGGAGCCTGATGGTGTGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTGCGGCTCTTGCCGAAATGAAGGGTATTCTCTACGCCAAACCTTGATCCGTGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGCTAAATGGCGGACAGACCCGTGCTGTCATTATTCGCCGTCCGACATTACT 900  
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysLaseArgIuThrGlyAlaGlnLeuGluIle 337  
RESULT 7  
ID ADC23791 standard; protein; 337 AA.  
XX  
AC ADC23791;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Protein sequence (SeqID 58) exhibiting nitrilase activity.  
XX  
KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO2003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR N-PSDB; ADC23790.  
XX  
PT Novel nitrilase polypeptide, useful for making (R) - or (S) -ethyl-4-cyano-

PT 3-hydroxybutyric acid or (R) - or (S) -mandelic acid or (S) - or (R) -phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 40; SEQ ID NO 58; 560pp; English.  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.  
XX  
SQ Sequence 337 AA;  
Alignment Scores:  
Pred. No.: 6.07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 7 Gaps: 1  
US-09-751-299-3 (1-1014) x ADC23791 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTCAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 VallysgluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGTGACAAACCATGAGTTGATGAGGAAGACAGCACGTAATATGCTGCTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATGCGCTTCCGAAACTTGATTCAGGCTACCATGGTTCTTGGCTTGACTGCACA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCATTGTAGCGCCAATACCATGAGAACTCATGTGAGTTGGATGCCCT 240  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTAGATGCAGCCCAAGCGGTGGAAATCATGTGCACCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGGTGACACCTTTACATCAGTCAGTGGTTCATAGCGATTAATGTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGGCGAAGTGAACCTACTTTGTGAACGTACTTTGTTGGCGGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTCGAGACGCTGTTGGAAGGCTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACACTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluIle 180  
QY 541 CATTGTGCGCTTGCCGAGCTTTAGCCTTTAATCCTTAATGCGGCGAAGCCCTGGGCGCT 600  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTACTAGCG 660

Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlnCysPheValLeuAla 220  
QY 661 TCGTGTGGCTGCTTTCACAAATCCATGCATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGGCTGTGTGTGACATCACTATCATAGGGCCGTGATGGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTCGCGCTCTTGCCGAAATGAAAGAGGTATCTCTACGCAAACTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlnGluAsnGlnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATGCGCGCAGACCCCTGCTGTCATTATCCCGCCGACATTACT 900  
Db 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlnHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATTTGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCCCACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
RESULT 8  
ADC23781  
ID ADC23781 standard; protein; 337 AA.  
AC ADC23781;  
XX 18-DEC-2003 (first entry)  
DT Protein sequence (SeqID 48) exhibiting nitrilase activity.  
DE  
XX  
KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO2003000840-A2.  
XX 03-JAN-2003.  
PD 15-MAY-2002; 2002WO-US015983.  
PF 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR N-PSDB; ADC23780.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 40; SEQ ID NO 48; 560bp; English.  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in

CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.

XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 6.07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 7 Gaps: 1

US-09-751-299-3 (1-1014) x ADC23781 (1-337)

QY 1 ATGAAAGAAGCTATCAAGGTGCGCTGCGTCAAGCCGCCCGATCTACATGATTGGAG 60  
Db 1 ValLysGlnAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGTGGACAAACCATTTGAGTTGAGAAAGACAGCAGCAGTAATGCTCGTCTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGlnAlaAlaArgAsnGlnYAlaArgLeu 40  
QY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTriPleProGlyTyrProTriPheLeuTriPheLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGSCCAATACATGAGAACTCATTTGATGGCCCT 240  
Db 61 AlaTriPlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGTAAAGCATTTCAAGATGACGCCAAGCGTTGGAAATCATGTCACCCCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGlnAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGCGTGGCACCCCTTACATCACTAGTGTTCATAGCGATAATGTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlnTriPleuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTAACGTACTTTGTCGGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGlnArgThrLeuPheGlyGlu 140  
QY 421 GGGATGTTTCAATCGCTAGCGGTTTTCAGACGTCGTGGAAGGCTGGGTGCTTATGC 480  
Db 141 GlysPglYserSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysCys 160  
QY 481 TGTGGAGACCTTCAACCGCTTAACAATAAGCTTTGTATGACAAATGAAGAGATT 540  
Db 161 CysTriPlyHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGlnGluIle 180  
QY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTATCCTAATGCGGCGAAAGCCTGGGGCT 600  
Db 181 HisPheAlaAlaTriProSerPheSerIleTyrArgGlnAlaThrGlnValLeuGlyPro 200  
QY 601 GATGTCAATGAGGCGCTCTGCAATCTATGCCGTGAAGGCGCAATGCTTCTACTAGCG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlnCysPheValLeuAla 220  
QY 661 TCGTGTGGCTGCTTTCACAAATCCATGCATGATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTGTCTTCTGGCTGTGTGTGACATCACTATCATAGGGCCGTGATGGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260



QY 781 GTGCGCCTTGCCGAAATGAGAGGGTATTCTTACGCCAAACCTTGATCCTGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATGCGCGAGACCCCTGTCATATTATCCCGTCGACATTACT 900  
Db 281 ArgIleTyrAlaIleThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTTAATTACCGGTAGTGAATTGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGTTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
RESULT 9  
ADH35882  
ID ADH35882 standard; protein; 337 AA.  
XX  
AC ADH35882;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SeqID48.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KM high throughput system; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI McQuaid J, Stege J;  
XX  
DR WPI; 2004-142708/14.  
DR N-PSDB; ADH35881.  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 48; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 6.07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6

DB: 8 Gaps: 1  
US-09-751-299-3 (1-1014) x ADH35882 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTCAAGCCGCCCGATCTACATGATTGGAG 60  
Db 1 VallysgluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGTGGCAAAACCATTGAGTTGATGGAAGAAGCAGCAGCAGTAATATCCTCGTCTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATCGCCTTTCCGAAACTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTGTGTACGCGCAATACATGAGAAGCACTCATTTGAGTTGGCCCT 240  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCCAAGCGGTTGGAAATCATGATGATCCCTGGGATG 300  
Db 81 GlnAlaIleArgIleSerGluAlaAlaGlnAlaArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGCGTGCCACCCCTTACATCACTCACTGAGTTCATAGGCGATAATGTTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTTGGCGGA 420  
Db 121 ThrAlaGlyLeuArgGlyLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGGAAGCGTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACACTTCAACCGCTAACAAATAACGCTTGTATGACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATGTGCGGCTTGCCCGAGCTTTAGCCTTATCCTAATGCGGCGAAAGCCCTGGGCCT 600  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCCCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTAACG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValAlaGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTGTTTCACAATTCATGATATGCTTGTATGACAGTGAAGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTGTCTTGCTGCTGTGTGAGACACTCAGTATCATAGGCGCTGATGGTGAAGT 780  
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGCGCCTTGTCGCCGAAATGAAGAGGTATTCTTACGCAAACTTGATCCTGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATGCGCGAGACCCCTGTCATATTATCCCGTCGACATTACT 900  
Db 281 ArgIleTyrAlaIleThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTTAATTACCGGTAGTGAATTGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGTTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
RESULT 10



ADH35892  
ID ADH35892 standard; protein; 337 AA.  
XX  
AC ADH35892;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SegID58.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI Mcquaid J, Stege J;  
XX  
DR WPI; 2004-142708/14.  
DR N-PSDB; ADH35891.  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 58; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 6.07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 8 Gaps: 1

US-09-751-299-3 (1-1014) x ADH35892 (1-337)

QY 1 ATGAAAGAGCTATCAAGCTGCGCTGCGCAAGCCGCCGATCTACATGATTTGGAG 60  
:::|||||  
DB 1 VallysGluAlaIleYsValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
|  
QY 61 GCGACGGTGGACAAACCATTTGATTTGATGGAAGACAGCAGTAATAATGCTGCTG 120  
|||||  
DB 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
|  
QY 121 ATCGCCTTTCGGAACCTGGATTCAGGCTACCCAGTCTTCTTGCTTGACTCACCA 180  
|||||  
DB 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
|  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCT 240

DB 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
|  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
|||||  
DB 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
|  
QY 301 AGTGAACGGGTGCGTGACCCCTTTACATCAGTCAGTGTTCATAGGCGATAATGTTGAC 360  
|||||  
DB 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
|  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTTACTTTGTTGAACGTACTTTGTTGGCGAA 420  
|||  
DB 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
|  
QY 421 GGGGATGCTTCATCGCTAGCGGTTTTCGAGACGTCGTGGAAGGCTGGGTGCTTATGC 480  
|||||  
DB 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysCys 160  
|  
QY 481 TGTGGGAGCACCTTCACACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATT 540  
|||||  
DB 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
|  
QY 541 CATTTGGCGGCTTGCCGAGCTTTAGCCTTTATCCTTAATGCGCGCAAGCCCTGGGCGCT 600  
|||  
DB 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
|  
QY 601 GATGTCATGTAGCGGCGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTGTACTAGCG 660  
:::|||||  
DB 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
|  
QY 661 TCGTGTGCGCTCGTTTCACCAATCATGATGCATATGCTTTGTACAGATGACGAAAGCAT 720  
|||||  
DB 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
|  
QY 721 GCGTTGCTTCTGCTGCTGTGTGACACTACGTCATCATAGGCGCTGATGCTGACTTG 780  
:::|||||  
DB 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
|  
QY 781 GTCGCGCCTCTTGCCGAAATGAAGAGGTAATCTCTACGCAACCTTGATCTGAGTA 840  
|||||  
DB 261 AlaArgProLeuGlyGluAsnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
|  
QY 841 CGCATCCTTGCTAAATGCGCGCAGACCTGCTGTCATTATCCGCTCCGACATTACT 900  
|||||  
DB 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
|  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAATGAAGTGATCTTCT 960  
|||||  
DB 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
|  
QY 961 CCTTACGCTTTGGTTAAAGCGTCTGAGACGGGTGCGCACTCGAA 1005  
320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
|  
RESULT 11  
ID ADG93583 standard; protein; 337 AA.  
XX  
AC ADG93583;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitrilase enzyme amino acid sequence SegID48.  
XX  
KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003097810-A2.  
XX  
PD 27-NOV-2003.



CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.

XX  
SQ Sequence 337 AA;

Alignment Scores:

Pred. No.:	6.07e-138	Length:	337
Score:	1339.00	Matches:	252
Percent Similarity:	84.2%	Conservative:	30
Best Local Similarity:	75.2%	Mismatches:	47
Query Match:	73.3%	Indels:	6
DB:	8	Gaps:	1

US-09-751-299-3 (1-1014) x ADG93593 (1-337)

QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTGCAGCCGCCCATCTACATGATTGGAG 60  
:::|||||  
Db 1 ValysgluAlaIlelySvalAcysValGlnAlaAlaProValPheLeuAspLeuasp 20  
QY 61 GCGACGCTGACAAACCATGAGTTGATGGAAGAAGCAGACGTAATATGCTGCTG 120  
|||||  
Db 21 AlaThrValAspIysThrValAlaLeuIleGlGlnAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATCGCCTTTCCGGAACCTTGATTCAGAGCTACCCATGTTCTTTGGCTTGACTCACCA 180  
|||||  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTGTGATCCGCCATATACATGAGAACTCATTTGGATGGCCCT 240  
|||||  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTACAGATGCAGCCAGCGGTTGGGAATCATGGTCAACCCTGGGATG 300  
|||||  
Db 81 GlnAlaIysArgIleSerGlnAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGGGTGGCACCCCTTACATCAGTCAGTGGTTCATAGGCGATAATGGTGAC 360  
|||||  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspIysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTTACTTTGTGAAACGTACTTTGTTGGCGAA 420  
|||||  
Db 121 ThrAlaGlyLeuAlaArgIysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGGTTCAATCGCTAGCGGTTTTCGAGACGTCTGTTGAAGGCTGGGCTTATGTC 480  
|||||  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATAAGAGATT 540  
|||||  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluIle 180  
QY 541 CATTTGCGGCTTGGCGGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCT 600  
|||||  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCATATGAGCGGCTCTGCAATCTATGCGGTTGAAGGCGCAATGCTTCTACTAGCG 660  
:::|||||  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTGTTTACAAATCCATGATGCATATGCTTGTACAGATGACGAAAGCAT 720  
|||||  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTGTCTTCTGCTGTGTGTGACACTACGATATGATAGGGCTGATGTTGTTGACTTG 780  
:::|||||  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTCCGGCCTCTTGCGGAATAATGAAGGGTATTCTTACGCCAAACCTTGATCCTGAGTA 840

Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGTAAATGCGCGAGACCCCTGCTGTCATTATTCCTCCGACATTACT 900  
|||||  
Db 281 ArgIleTyrAlaIysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCT 960  
|||||  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
RESULT 13  
AD162180  
ID AD162180 standard; protein, 337 AA.  
XX AC AD162180;  
XX AC 22-APR-2004 (first entry)  
XX DT  
XX DE Nitrilase polypeptide #24.  
XX KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipaeamic; enzyme.  
XX OS Unidentified.  
XX PN WO2003106415-A2.  
XX PD 24-DEC-2003.  
XX PF 13-JUN-2003; 2003WO-US018840.  
XX PR 13-JUN-2002; 2002US-0389317P.  
XX PR 28-JUN-2002; 2002US-0392944P.  
XX PA (DIVE-) DIVERSA CORP.  
XX PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX DR WPI; 2004-090821/09.  
XX DR N-PSDB; AD162179.  
XX PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX PS Claim 41; SEQ ID NO 48; 253bp; English.  
XX XX The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrilase polypeptide obtained from an  
XX environmental sample.





DB:	8	Gaps:	1
US-09-751-299-3 (1-1014) x ADI62190 (1-337)			
OY	1	ATGAAGAAGCTATCAAGGTCGCCTGCGTGAAGCCCGCATCTACATGATTTGGAG	60
Db	1	ValLysGluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp	20
OY	61	GCGACGGTGGACAAAACCATTTGAGTTGATGGAAGAACGACGACGTAATATGCTCTGTG	120
Db	21	AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu	40
OY	121	ATCGCCTTCCGGAACCTTGATTCGAGGCTAACCCATGCTTCTTGGCTTGACTCACCA	180
Db	41	IleAlaPheProGluThrTrpIleProGlyIleProThrPheLeuTrpLeuAspSerPro	60
OY	181	GCATGGGCAATGCAATTTGTACGCCAATATCCATGAGAATCTCATTTGAGTTGGACCT	240
Db	61	AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro	80
OY	241	CAAGCTAAGCGCATTTACAGATGACGCCAAGCGGTTGGAAATCATGTCACCTGGGATG	300
Db	81	GlnAlaLysArgLieserGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr	100
OY	301	AGTGAACGGGTGGTGGACACCTTTACATCATGTCAGTGTTCATAGCGCATTAATGTTGAC	360
Db	101	SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu	120
OY	361	ACCATTTGGGGCCCGCGAAGTGAACCTACTTTTGTGAACGTACTTTGTTCCGGCGAA	420
Db	121	ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu	140
OY	421	GGGATGTTCAATCGCTAGCGGTTTTCGAGACGTCGTGGAAGGCTGGGCTTATGC	480
Db	141	GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysLeuCys	160
OY	481	TGTTGGAGCACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATT	540
Db	161	CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle	180
OY	541	CATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCT	600
Db	181	HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro	200
OY	601	GATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGCAATGCTTCGTAAGCG	660
Db	201	GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla	220
OY	661	TGCTGTGCGCTGTTTACAATCCATGATCGATGCTTGTACAGATGACGAAAGCAT	720
Db	221	SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis	240
OY	721	GCGTTGCTTCTGCTGGTGGTGGACACTCACGTATCATAGGGCGCTGATGGTGACTTG	780
Db	241	SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu	260
OY	781	GTCGGCCTCTTGCCGAAATGAAGAGGATTTCTTACGCCAAACCTTGATCCTGGAATA	840
Db	261	AlaArgProLeuGlyGluAsnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla	280
OY	841	CGCATCCTTGCTAAATGGCGGAGACCCCTGCTGTCATTAATCCCGTCCCGACATTA	900
Db	281	ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr	300
OY	901	CGCTTGCTAATAGATCGACGCCCTAAATTACCGGTAGTGAATGAAGGTGATCTTGT	960
Db	301	ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle---	319
OY	961	CCTTACGCTTTGGTAAAGCTCTGAGACGGGTGCGCAACTCGAA	1005
Db	320	-----ProAlaSerAlaGlnGlyPheGluValGlu	329

RESULT 15

ADI64311	AD I64311 standard; protein; 337 AA.
XX	AC ADI64311;
XX	DT 22-APR-2004 (first entry)
XX	DE Nitrilase seq id 30.
XX	KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
KW	carboxylic acid; cyanohydrin moiety hydrolysis;
KW	aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
KW	chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
KW	(R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
KW	(R)-phenyl lactic acid derivative; & enantiomeric excess;
KW	& diastomeric excess; food additive; drug intermediate; nitrilase.
XX	Unidentified.
OS	US2004014195-A1.
XX	PN 22-JAN-2004.
XX	PD 15-MAY-2003; 2003US-00440523.
XX	PF 29-DEC-1999; 99US-0173609P.
XX	PR 07-DEC-2000; 2000US-0254414P.
PR	28-DEC-2000; 2000US-00751299.
PR	21-JUN-2001; 2001US-0300189P.
PR	30-JUL-2001; 2001US-0309006P.
PR	22-JAN-2002; 2002US-0351336P.
PR	15-MAY-2002; 2002US-00146772.
PR	09-SEP-2002; 2002US-00241742.
XX	(DIVE-) DIVERSA CORP.
PA	Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
PI	WPI, 2004-121569/12.
XX	DR N-PSDB; ADI64310.
XX	Novel isolated or recombinant polypeptide having nitrilase activity,
PT	useful in production of food additives.
PT	Claim 46; SEQ ID NO 58; 105bp; English.
PS	The invention describes an isolated or recombinant polypeptide (I)
XX	comprising amino acids having a sequence at least 50 % identical to a
CC	sequence (S1) available in electronic form (EC) from the following web
CC	site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its
CC	variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC	residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC	fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC	hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC	hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
CC	chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC	producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC	mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
CC	derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
CC	and for identifying a modified compound. The inventive method is useful
CC	for monitoring or determining & enantiomeric excess or & diastomeric
CC	excess. (I) is useful in the production of food additives and drug
CC	intermediates. This is the amino acid sequence of a nitrilase of the
CC	invention.
XX	Sequence 337 AA;
SO	Alignment Scores:
SO	Pred. No.: 6.07e-138 Length: 337
SO	Score: 1339.00 Matches: 252
SO	Percent Similarity: 84.2% Conservative: 30
SO	Best Local Similarity: 75.2% Mismatches: 47
SO	Query Match: 73.3% Indels: 6



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:35:46 ; Search time 7.1054 Seconds  
(without alignments)  
2359.702 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826

Sequence: 1 atgaaagaagctatcaaggt.....cgcaactcgagaatttga 1014

Scoring table:

BL0SUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 segs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q=/abs/ABSSWEB\_spool/US09751299/runat 26042006 090039 18719/app query.fasta\_1  
-DB=Issued Patents\_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bite -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abs02p -USER=US09751299 @CCG 1 1 101 @runat 26042006 090039 18719  
-NCPU=6 -ICPU=3 -NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THEADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_AA:\*

1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	955.5	52.3	344	1	US-08-690-493-1 Sequence 1, Appli
2	822	45.0	356	2	US-09-806-876A-2 Sequence 2, Appli
3	776	42.5	369	2	US-09-823-373-5 Sequence 5, Appli
4	776	42.5	369	2	US-09-823-373-14 Sequence 14, Appli
5	638	34.9	354	1	US-08-447-702-5 Sequence 5, Appli
6	638	34.9	354	1	US-08-465-615-5 Sequence 5, Appli
7	152	8.3	267	2	US-09-902-540-14601 Sequence 14601, A
8	139.5	7.6	315	2	US-10-105-294B-2 Sequence 2, Appli
9	139	7.6	311	1	US-07-917-111-5 Sequence 5, Appli
10	139	7.6	311	1	US-07-917-111-6 Sequence 6, Appli
11	139	7.6	311	1	US-08-479-638-5 Sequence 5, Appli
12	139	7.6	311	1	US-08-479-638-6 Sequence 6, Appli

13	139	7.6	311	1	US-08-294-871A-70	Sequence 70, Appli
14	139	7.6	311	2	US-08-876-398A-70	Sequence 70, Appli
15	133	7.3	303	1	US-08-294-871A-46	Sequence 46, Appli
16	133	7.3	303	2	US-08-876-398A-46	Sequence 46, Appli
17	132	7.2	303	1	US-08-294-871A-36	Sequence 36, Appli
18	132	7.2	303	2	US-08-876-398A-36	Sequence 36, Appli
19	131	7.2	303	1	US-08-294-871A-4	Sequence 4, Appli
20	131	7.2	303	1	US-08-294-871A-16	Sequence 16, Appli
21	131	7.2	303	1	US-08-294-871A-34	Sequence 34, Appli
22	131	7.2	303	1	US-08-294-871A-48	Sequence 48, Appli
23	131	7.2	303	1	US-08-294-871A-58	Sequence 58, Appli
24	131	7.2	303	1	US-08-294-871A-68	Sequence 68, Appli
25	131	7.2	303	2	US-08-876-398A-4	Sequence 4, Appli
26	131	7.2	303	2	US-08-876-398A-16	Sequence 16, Appli
27	131	7.2	303	2	US-08-876-398A-34	Sequence 34, Appli
28	131	7.2	303	2	US-08-876-398A-48	Sequence 48, Appli
29	131	7.2	303	2	US-08-876-398A-58	Sequence 58, Appli
30	131	7.2	303	2	US-08-876-398A-68	Sequence 68, Appli
31	130	7.1	303	1	US-08-294-871A-38	Sequence 38, Appli
32	130	7.1	303	1	US-08-294-871A-44	Sequence 44, Appli
33	130	7.1	303	1	US-08-294-871A-50	Sequence 50, Appli
34	130	7.1	303	1	US-08-294-871A-56	Sequence 56, Appli
35	130	7.1	303	1	US-08-294-871A-60	Sequence 60, Appli
36	130	7.1	303	2	US-08-876-398A-38	Sequence 38, Appli
37	130	7.1	303	2	US-08-876-398A-44	Sequence 44, Appli
38	130	7.1	303	2	US-08-876-398A-50	Sequence 50, Appli
39	130	7.1	303	2	US-08-876-398A-56	Sequence 56, Appli
40	130	7.1	303	2	US-08-876-398A-60	Sequence 60, Appli
41	130	7.1	304	1	US-08-900-711-5	Sequence 5, Appli
42	130	7.1	304	1	US-08-815-356-2	Sequence 2, Appli
43	130	7.1	304	1	US-08-415-343B-5	Sequence 5, Appli
44	130	7.1	304	1	US-08-762-433-1	Sequence 1, Appli
45	130	7.1	304	1	US-09-001-219-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-1  
; Sequence 1, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000e1 Nitriase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acid residues



; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Gordona terrae
; STRAIN: MA-1
; CELL TYPE: unicellular organism
; FEATURE:
; OTHER INFORMATION: Xaa is Met or a deletion
;
US-08-690-493-1

Alignment Scores:
Pred. No.: 7.98e-97 Length: 344
Score: 955.50 Matches: 182
Percent Similarity: 73.4% Conservative: 47
Best Local Similarity: 58.3% Mismatches: 82
Query Match: 52.3% Indels: 1
DB: 1 Gaps: 1

US-09-751-299-3 (1-1014) x US-08-690-493-1 (1-344)

QY 16 AAGTGCCTGCGTCAAGCCGCCGATCTACATGATTGGAGGCGACGGTGACAA 75
Db 10 LysAlaAlaValThrGlnAlaGluProValTrpPheAspLeuSerAlaThrValAspLys 29
QY 76 ACCATTGAGTTGATGAAGAAGCAGCAGCTAATATGCTCTGATCGCCCTTCGGAA 135
Db 30 ThrIleAlaLeuValGluGluAlaSerArgAlaGlyAlaAspLeuIleAlaPheProGlu 49
QY 136 ACTTGATTCAGGCTACCATGGTTCTTTGGCTTGACTCACCATGGGCAATGCAA 195
Db 50 ThrTrpIleProGlyTyrProGlnAsnSerLeuTrpLeuAspSerValAlaTrpGlnSerGln 69
QY 196 TTTGTACGCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCTCAAGCTAAGCGCATT 255
Db 70 TyrPheIleArgTyrProGlnAsnSerLeuAspLeuAspGlySerGluPheAlaIle 89
QY 256 TCAGATGACGCCAAGCGGTTGGGAATCATGTGTACCCCTGGGGATGATGAACGGGTGGT 315
Db 90 ArgGluAlaAlaArgLysAsnAspIleAlaIleThrMetCglyPheSerGluArgGlyHis 109
QY 316 GGCAACCTTTACATCAGTCAGTGGTTTCATAGGCGATAAGGTGACACCATTTGGGCCCGG 375
Db 110 GlySerLeuTyrMetGlyGlnAlaValIleGluArgAspGlyValValValArgThrArg 129
QY 376 CGAAAGTTGAAACCTACTTTTGTGTAACGTACTTTGTCGGCGAAGGGGATGTTCAATCG 435
Db 130 ArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGluGlyAspGlySerAsp 149
QY 436 CTAGCGGTTTTCGAGACGCTGTGTTGGAAGGCTGGGTGGCTTAATGCTGTGGAGCACCTT 495
Db 150 LeuValValAspGlnThrSerLeuGlyArgValGlySerLeuCysCysTrpGluHisLeu 169
QY 496 CAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGATTCATTGTGCGGCTTGG 555
Db 170 GlnProLeuThrLysThrAlaMetCtyrSerGlnHisGlnGlnIleHisIleAlaIleTrp 189
QY 556 CCGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAGCG 615
Db 190 ProSerPheSerIlePheProGlyAlaValTyrAlaLeuGlyProGluValAsnThrAla 209
QY 616 GCCTTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTACTAGCGTGTGCGCTGCT 675
Db 210 AlaSerGlnGlnTyrAlaValGluGlnGlnThrTyrValLeuAlaProCysAlaValIle 229
QY 676 TCACAATCCATGATGATATGCTTTGTACAGATGACGAAGAAGCATGCGTTGCTTGCGT 735
Db 230 GlyAspAlaGlyTyrGluAlaPheAlaAspThrGluGluLysArgGlnLeuIleHisLys 249
QY 736 GGTGGTGACACTCAGCTATCATAGGGCTGATGTTGTTGACTTGTCGGCCTCTTGGC 795
Db 250 GlyGlyGlyTyrAlaArgIleTyrCglyProAspGlyArgSerLeuAlaGluProLeuAla 269

QY 796 GAAATGAGAGGGTATTCTCTACGCCAACTTGATCCGTGAGTACGCATCCTTGCTAA 855
Db 270 ProAsnAspGluGlyIleLeuTyrAlaAspIleAspLeuSerAlaIleLeuAlaLys 289
QY 856 ATGGCGGCAAGCCCTGCTGTCAATTATCCCGTCCGACATTAATGCTTGTGCTAATAGAT 915
Db 290 AsnProAlaAspProValGlyHisTyrSerArgProAspValLeuArgLeuGlyPheAsn 309
QY 916 CGCAGCCCTAAATTACCGGTAGTTGAATTGAAGT 951
Db 310 LysAlaProGln---ProLysValaAsnIleLeuGly 320

RESULT 2
US-09-806-876A-2
; Sequence 2, Application US/09806876A
; Patent No. 6869783
; GENERAL INFORMATION:
; APPLICANT: Ress-Loeschke, Marion
; APPLICANT: Friedrich, Thomas
; APPLICANT: Hauer, Bernhard
; TITLE OF INVENTION: A process for preparing chiral carboxylic acids
; TITLE OF INVENTION: from nitriles using a nitrilase or microorganisms
; TITLE OF INVENTION: which comprise a gene for the nitrilase
; FILE REFERENCE: 49462
; CURRENT APPLICATION NUMBER: US/09/806, 876A
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: Germany/19848129.2
; PRIOR FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 2
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Alcaligenes faecalis
;
US-09-806-876A-2

Alignment Scores:
Pred. No.: 4.87e-82 Length: 356
Score: 822.00 Matches: 160
Percent Similarity: 65.1% Conservative: 56
Best Local Similarity: 48.2% Mismatches: 114
Query Match: 45.0% Indels: 2
DB: 2 Gaps: 1

US-09-751-299-3 (1-1014) x US-09-806-876A-2 (1-356)

QY 4 AAAGAAGCTATCAAGTTCGCCCTGCGTCAAGCCGCCGATCTACATGATTGGAGCG 63
Db 4 ArgLysIleValArgAlaAlaAlaValGlnAlaAlaSerProAsnTyrAspLeuAlaThr 23
QY 64 ACGGTGACAAACCATTTGAGTTGATGGAAGAGCAGACAGTAATATGCTGCTGTGATC 123
Db 24 GlyValAspLysThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCysAspLeuIle 43
QY 124 GCCTTCCGAAACTTGATTCAGGCTACCGTACGTTCTTTGGCTTGACTCACCGCA 183
Db 44 ValPheGlyGluThrTrpLeuProGlyTyrProPheHisValTrpLeuGlyAlaProAla 63
QY 184 TGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGAGTGGCCCTCAA 243
Db 64 TrpSerLeuLysTyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGlu 83
QY 244 GCTAAGCGCATTTGAGATGACGCCAAGCGGTTGGGAATCATGTGTCACCCGTGGATGAGT 303
Db 84 PheGlnArgIleAlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSer 103
QY 304 GAACGGGTGCGTGACCCCTTTACATCAGTCAGTGGTTTCATAGGCGATAATGTTGACACC 363
Db 104 GluArgSerGlyGlySerLeuTyrLeuGlyGlnCysLeuIleAspAspLysGlyGluMet 123
QY 364 ATTGGGGCCCGGCAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGG 423
Db 124 LeuTrpSerArgArgLysLeuLysProThrHisValGluArgThrValPheGlyGluGly 143

QY 424 GATGGTTCATCGCTAGCGGTTTTCGAGACGTCTGTGGAAGGCTGGGTGCTTATGCTGT 483  
Db 144 TyrAlaArgAspLeuIleValSerAspThrGluLeuGlyArgValGlyAlaLeuCysCys 163  
QY 484 TGGGAGCACCTTCAACCCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATTTCAT 543  
Db 164 TrpGluHisLeuSerProLeuSerLysTyrAlaLeuTyrSerGlnHisGluAlaIleHis 183  
QY 544 TGTGGCGCTTGGCCGAGCTTTAGCCTTTATCCTTAATGCGCGAAAGCCCTGGGCGCTGAT 603  
Db 184 IleAlaIleTrpProSerPheSerLeuTyrSerGluGlnAlaHisAlaLeuSerAlaLys 203  
QY 604 GTCAATGTAGCGGCTCTGCAATCTATGCGCTTGAAGGCAATGCTTCGTACTAGCGTCG 663  
Db 204 ValAsnMetAlaAlaSerGlnIleTyrSerValGluGlyGlnCysPheThrIleAlaIle 223  
QY 664 TGTGGCGCTGTTTCAACAATCGATGATGCTTGTGTACAGATGACGAAAGCATGCG 723  
Db 224 SerSerValValThrGlnGluThrLeuAspMetLeuGluValGlyGluHisAsnAlaPro 243  
QY 724 TTGCTTCTGGCTGGTGGTGAACATCAGTATCATAGGCGCTGATGCTGCTGCTGCTG 783  
Db 244 LeuLeuLysValGlyGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAla 263  
QY 784 GCGGCTCTTGCCGAAATGAAGGGTATTCTCTACGCAACCTTGATCCTGAGTACGC 843  
Db 264 ProTyrLeuProHisAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluIle 283  
QY 844 ATCCTTGTAAATGGCGGAGACCCCTGCTGATTAATCCCGTCCGACATTAATCTGC 903  
Db 284 AlaPheAlaLysAlaIleAsnAspProValGlyHisTyrSerLysProGluAlaThrArg 303  
QY 904 TTGCTAATAGATCGCAGCCCTTAATTAACCGTAGTGAATTGAAGTGATCTTCCTCT 963  
Db 304 LeuValLeuAspLeuGlyHisArgAspPrometThrArgValHisSer-----LysSer 321  
QY 964 TACGCTTTGGGTAAGCGCTCTGAGACGGGTCCGCA 999  
Db 322 ValThrArgGluGluAlaProGluGlnGlyValGln 333

RESULT 3  
US-09-823-373-5  
; Sequence 5, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; TITLE OF INVENTION: Acidovorax Facilis 72W  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 5  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Acidovorax Facilis  
US-09-823-373-5

Alignment Scores:  
Pred. No.: 6.12e-77  
Score: 776.00  
Percent Similarity: 67.04  
Best Local Similarity: 47.64  
Query Match: 42.54  
DB: 2

Length: 369  
Matches: 150  
Conservative: 61  
Mismatch: 100  
Indels: 4  
Gaps: 3

US-09-751-299-3 (1-1014) x US-09-823-373-5 (1-369)  
QY 22 GCCTGCGTCAAGCCCGCCGATCTACATGATTTGGAGGCGACCGGTGCAAAACCAT 81  
Db 11 AlaThrValGlnAlaGluProValTyrLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGAAGAAGCAGCAGCATTAATGCTGCTGATGCTCTTCCGAAACTTGG 141  
Db 31 GlyIleIleGluGlnAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCCAGGCTACCCATGTTCTTGTGCTGACTCACACGATGGGCAATGCAATTGTA 201  
Db 51 IleProGlyTyrProTyrTrpAlaTyrLeuGlyAspValLysTyrSerLeuSerPheThr 70  
QY 202 CGCCATACCATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTTAAGCCATTTCAGAT 261  
Db 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCAGCGGTTGGGAATCATGCTCACCCCTGGGATGATGACCGGTCGTCACACC 321  
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
QY 322 CTTTACATCAGTCAAGTGTTCATAGGCGAATATGCTGACACACCATTTGGGCGGAAAG 381  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGAACGTAATTTGTCGGGAGGAGATGTTTCATCGCTAGCG 441  
Db 131 LeuLysProThrHisValGluArgThrIleTyrGlyGlyGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTCGAGAGCTGTGGAAGGCTGGGTGCTTATGCTGTTGGAGACCTTCAACCG 501  
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCystrProGluHisPheGlnPro 170  
QY 502 CTACAAATAACGCTTTGTATGACCAAAATGAAGAATTCATTGTGCGGCTTGGCCGAGC 561  
Db 171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTrpProAla 190  
QY 562 TTTAGC--CTTATCTTAATGCGGCGAAAGCCCTGGGCGCTGATGATGATGAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTGAATCTATGCGGCTTGAAGGCGAAATGCTTCTGACTAGCGTGTGCGCTGTTCA 678  
Db 210 ThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCTTCTGCTGCTGCT 738  
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249  
QY 739 GGTGACACTCAGTATCATAGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798  
Db 250 CysGlyTyrPheAlaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269  
QY 799 AATGAAGGATATTCTTACGCAAACTTGATCCTGAGTACGCAATCTTGTCTAAATG 858  
Db 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuLeuAlaLysAla 289  
QY 859 GCGGAGACCGCTGCTGCTATTAATCCCGTCCGACATTAATCTGCTGCTGCTGCTGCTG 918  
Db 290 GlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309  
QY 919 AGCCCTAAATTAACGGTA-----GTTGAATTTGAAGTGATGTT 957  
Db 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 4  
US-09-823-373-14  
; Sequence 14, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita

```

; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Acidovorax facilis
; US-09-823-373-14
```

## Alignment Scores:

Pred. No.:	6.12e-77	Length:	369
Score:	776.00	Matches:	150
Percent Similarity:	67.0%	Conservative:	61
Best Local Similarity:	47.6%	Mismatches:	100
Query Match:	42.5%	Indels:	4
DB:	2	Gaps:	3

US-09-751-299-3 (1-1014) x US-09-823-373-14 (1-369)

```

QY 22 GCCTGCGTGCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGACAAACCATT 81
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11 AlathrValGlnAlaGluProValTrpLeuAspAlaThrIleAspIleSerIle 30
QY 82 GAGTTGATGGAAGAAGACGACGACGTAATATGCTGCTGATCGCTTCCGAAACTTGG 141
   ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 31 GlyIleIleGluGluAlaIleAlaGlnIleGlyAlaSerIleIleAlaPheProGluValPhe 50
QY 142 ATTCAGGCTACCCATGGTCTTGGCTTGACTCACGACATGGGCAATGCAATTGTA 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51 IleProGlyTyrProTyrTrpAlaTrpLeuGlyAspValIleTyrSerLeuSerPheThr 70
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCATTTAGAT 261
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 SerArgTyrHisGlnAsnSerLeuGluLeuGlyAspAspArgMetArgLeuGlnLeu 90
QY 262 GCAGCCCAAGCGGTTGGGAATCATGCTACCCCTGGGGATGATGAACGGGTCGGTGCAAC 321
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110
QY 322 CTTTACATCAGTCAGTGGTTTCATAGCGGATAATGTTGACACATTGGGGCCCGCGCAAG 381
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130
QY 382 TTGAACCTACTTTTGTGTAACGTACTTTGTTGGCGGAAGGATGTTTCATCGCTAGCG 441
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 LeuLysProThrHisValGlnArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150
QY 442 GTTTTCGAGACGTCTGTTGGAAGCTGGGTGGCTTATGCTGTTGGAGACACCTTCAACG 501
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCystrpGluHisPheGlnPro 170
QY 502 CTAACAAATAACGCTTTGTATGACAAATAAGAGATTCATTGTGCGGCTTGCGCGGAGC 561
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
Db 171 LeuSerLysPheMetMetLysSerLeuGlyGluGlnValHisValAlaSerTrpProAla 190
QY 562 TTTAGC---CTTATCCTAATGCGGCAAGCCCTGGGCGCTGATGTCAATGTAGCGGCC 618
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209
QY 619 TCTGAATCTATGCCGTTGAAGGCAATGCTTCTGACTAGCGTGTGCGCTGCTTCA 678
   ::||| ||||| ::||| ||||| ||||| ||||| ||||| |||||
Db 210 ThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCySerThrGlnValIleGly 229
```

```

QY 679 CAATCCATGATCATATGCTTTGTACAGATGACGAAAGCATGCCGTCTTGCGTGGT 738
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 ProSerAlaIleGluThrPheCybLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249
QY 739 GGTGACACTACGATCATAGGCGCTGATGTGGTGACTTGGTCCGCTTTGCCGAA 798
   ||| ::||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 CybGlyTrpAlaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269
QY 799 AATGAAGGGTATTTCTTACGCAACCTTGATCCTGAGTACGCATCCTTGCTAAATG 858
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuLeuAlaLysAla 289
QY 859 GCGGCAAGCCCTGCTGTGATTAATCCCGTCCCGACATTACTCGCTTGAATAGATCGC 918
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 290 GlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309
QY 919 AGCCCTAAATTACCGGTA-----GTGAATTGAAAGTGATCTT 957
   ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324
```

## RESULT 5

US-08-447-702-5

Sequence 5, Application US/08447702

Patent No. 5629190

GENERAL INFORMATION:

APPLICANT: Petre, Dominique

APPLICANT: Cerbeleaud, Edith

APPLICANT: Levy-Schil, Sophie

APPLICANT: Crouzet, Joel

TITLE OF INVENTION: Polypeptides Possessing A Nitrilase

TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,

Patent No. 5629190

TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them

TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy-

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER &amp; MATHIS

STREET: P.O. Box 1404

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/447,702

FILING DATE: 23-MAY-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/194,588

FILING DATE: 10-FEB-1994

PRIOR APPLICATION DATA: FR 92-09-882

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 003025-019

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6620

TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 354 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-447-702-5

Alignment Scores:  
Pred. No.: 1.14e-61 Length: 354  
Score: 638.00 Matches: 134  
Percent Similarity: 60.3% Conservative: 62  
Best Local Similarity: 41.2% Mismatches: 107  
Query Match: 34.9% Indels: 22  
DB: 1 Gaps: 5

US-09-751-299-3 (1-1014) x US-08-447-702-5 (1-354)

```
QY 13 ATCAAGTCGCCCTGCGTCAAGCCGCCGATCTACATGATTGGAGCGGACGGTGAC 72
Db 7 VallysValAlaAlaValGlnAlaAlaProValPheMetAsnLeuGluAlaThrValAsp 26
QY 73 AAACCATGAGTTGATGGAAGACGACGCTAATAATGCTGCTGATCGCCTTCCG 132
Db 27 LysThrCysLysLeuIleAlaGluAlaAlaSerMetGlyAlaLysValIleGlyPhePro 46
QY 133 GAAACTTGATTCAGGCTACCCATGGTTTCTTGGCTTGACTCA----- 177
Db 47 GluAlaPheIleProGlyTyrProTyrTrpIleTrpThrSerAsnMetAspPheThrGly 66
QY 178 CCAGCATGGGCAATGCAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGC 237
Db 67 MetMetTrpAlaValLeuPhe-----LysAsnAlaIleGluIleProSer 81
QY 238 CCTCAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGGAATCATGTACCCCTGGGG 297
Db 82 LysGluValGlnGlnIleSerAspAlaAlaLysLysAsnGlyValTyrValCysValSer 101
QY 298 ATGAGTGAACGGGTGCGTGCGACCCCTTACATCAGTCAGTGGTTCATAGGCGATAATGCT 357
Db 102 ValSerGluLysAspAsnAlaSerLeuTyrLeuThrGlnLeuTrpPheAspProAsnGly 121
QY 358 GACACCATTTGGGCGCCGCGAAAGTTGAAACCTACTTTGTTGAACGACTTTGTTGGC 417
Db 122 AsnLeuIleGlyLysHisArgLysPheLysProThrSerSerGluArgAlaValTrpGly 141
QY 418 GAAGGGGATGTTTCATCGCTACGCGTTTTCGAGACGCTGTTGGAAGGCTGGGTGCTTA 477
Db 142 AspGlyAspGlySerMetAlaProValPheLysThrGluTyrGlyAsnLeuGlyGlyLeu 161
QY 478 TGCTGTTGGAGCACCTTCAACCGCTTAACAAATACGCTTTGTATGCACAAATGAAGAG 537
Db 162 GlnCysTrpGluHisAlaLeuProLeuAsnIleAlaAlaMetGlySerLeuAsnGluGln 181
QY 538 ATTCAATTGCGCGCTTGCGCGAGCTTTAGCCTTTAATCTTAATGCGCGAAAGCCCTGGGG 597
Db 182 ValHisValAlaSerTrpProAlaPhe-----ValProLysGlyAlaValAlaSerSerArg 199
QY 598 CCTGATGTCAATGTAGCGGC-----TCTGGAATCTAAT 630
Db 200 ValSerSerSerValCysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyr 219
QY 631 GCCGTTGAAGGCAATGCTTCGTACTAGCGTGTGCGCTGCTTCAATCCATCATCATC 690
Db 220 AlaIleSerAsnGlnValTyrValIleMetSerThrAsnLeuValGlyGlnAspMetIle 239
QY 691 GATATGCTTTGTACAGATGACGAAAGCATGCGTTGCTTCTGCTGTGTGTGACACTCA 750
Db 240 AspMetIleGlyLysAspGluPheSerLysAsnPheLeuProLeuGlySerGlyAsnThr 259
QY 751 CGTATCATAGGGCGCTGATGGTGTGACTTGTGCGCGCTCTTGCCGAAATGAAGAGGT 810
Db 260 AlaIleIleSer--AsnThrGlyGluIleLeuAlaSerIleProGlnAspAlaGluGly 278
QY 811 ATTCTTACGCAAACTGATCTTGAGATAGCATCCTTGCTAAATGCGCGACAGCCCT 870
Db 279 IleAlaValAlaGluIleAspLeuAsnGlnIleIleTyrGlyLysTrpLeuLeuAspPro 298
QY 871 GCTGTCATTATTCGCCGTCGCCGACATTACTCGCTTGCTAATAGATCGAGCCCTAAATTA 930
Db 299 AlaGlyHisTyrSerThrProGlyPheLeuSerLeuThrPheAspGlnSerGluHisVal 318
```

```
QY 931 CCGTAGTTGAATT 945
Db 319 ProValLysLysIle 323
```

```
RESULT 6
US-08-465-615-5
; Sequence 5, Application US/08465615
; Patent No. 5635391
; GENERAL INFORMATION:
; APPLICANT: PETRE, Dominique
; APPLICANT: CERBELEAUD, Edith
; APPLICANT: LEVY-SCHIL, Sophie
; APPLICANT: CROUZET, Joel
; TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE
; TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,
; TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM
; TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO
; TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,615
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,588
; FILING DATE: 10-FEB-1994
; APPLICATION NUMBER: FR 9209882
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Crane-Feury, Sharon E
; REGISTRATION NUMBER: 36,113
; REFERENCE/DOCKET NUMBER: 003025-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-465-615-5
```

Alignment Scores:  
Pred. No.: 1.14e-61 Length: 354  
Score: 638.00 Matches: 134  
Percent Similarity: 60.3% Conservative: 62  
Best Local Similarity: 41.2% Mismatches: 107  
Query Match: 34.9% Indels: 22  
DB: 1 Gaps: 5

US-09-751-299-3 (1-1014) x US-08-465-615-5 (1-354)

```
QY 13 ATCAAGTCGCCCTGCGTCAAGCCGCCGATCTACATGATTGGAGCGGACGGTGAC 72
Db 7 VallysValAlaAlaValGlnAlaAlaProValPheMetAsnLeuGluAlaThrValAsp 26
QY 73 AAACCATGAGTTGATGGAAGACGACGCTAATAATGCTGCTGATCGCCTTCCG 132
Db 27 LysThrCysLysLeuIleAlaGluAlaAlaSerMetGlyAlaLysValIleGlyPhePro 46
```





Db 232 ThrAlaArgAlaSerGluGlyGluGlyLeuAlaLeuAlaProValAspPro 248

RESULT 8

US-10-105-294B-2

; Sequence 2, Application US/10105294B

; Patent No. 6800464

; GENERAL INFORMATION:

; APPLICANT: DRAUZ, KARLHEINZ

; APPLICANT: MAY, OLIVER

; APPLICANT: BOMMARIUS, ANDREAS

; APPLICANT: SYLDATK, CHRISTOPH

; APPLICANT: ALTENBUCHNER, JOSEF

; APPLICANT: WERNER, MARKUS

; APPLICANT: SIEMANN-HERZBERG, MARTIN

; TITLE OF INVENTION: D-Carbamoylase from Arthrobacter crystallopoietes

; FILE REFERENCE: 220024US0X

; CURRENT APPLICATION NUMBER: US/10/105,294B

; CURRENT FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: DE 101 14 999.9

; PRIOR FILING DATE: 2001-03-26

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2

; LENGTH: 315

; TYPE: PRT

; ORGANISM: ARTIFICIAL SEQUENCE

; FEATURE:

; OTHER INFORMATION: SYNTHETIC DNA

US-10-105-294B-2

Alignment Scores:

Pred. No.:	1.65e-06	Length:	315
Score:	139.50	Matches:	77
Percent Similarity:	39.1%	Conservative:	47
Best Local Similarity:	24.3%	Mismatches:	134
Query Match:	7.6%	Indels:	59
DB:	2	Gaps:	16

US-09-751-299-3 (1-1014) x US-10-105-294B-2 (1-315)

QY 67 GTGGACAAACCATTGAGTTGATGGAAGAAGCAGCAGCACTAATATGCTCGTTCGATCGCC 126

Db 25 ValAlaArgLeuIleAlaLeuGluGluValAlaAlaSerGlnGlyAlaGluLeuValAl 44

QY 127 TTTCGGGAA-----ACTTGATTCAGAGCTACCCATGCTTCTTGGCTTGAC 174

Db 45 PheProGluLeuThrLeuThrPhePhePro--ArgThrTrpPheGluGluGlyAsp 63

QY 175 TCACCAGCATGGCAATGCAATTGTACGCCAATACCATGAGAATCATTCATTGAGTTGAT 234

Db 64 -----PheGluGluTrpPheAspLysSerMetProAsnAsp 75

QY 235 GGCCCTCAAGCTAAGCCGATTTCAGATGCAAGCCGAGGTTGGGAATCATGTGATCCCTG 294

Db 76 -----AspValAlaProLeuPheGluArgAlaLysAspLeuGlyValGlyPheTyrLeu 93

QY 295 GGGATGAGTGAACGGGTCGTGGC--ACCCCTTACATCAGTCAGTGTTCATTAGCGCAT 351

Db 94 GlyTyrAlaGluLeuThrSerAspGluLysArgTyrAsnThrSerIleLeuValAsnLys 113

QY 352 AATGGTGACACCATTTGGCCCGCGGAAG----- 381

Db 114 HisGlyAspIleValGlyLysTyrArgLysMetHisLeuProGlyHisAlaAspAsnArg 133

QY 382 -----TTGAACCTACTTTGTTGAACGTACTTTGTTGGCGAAGGGGATGTTTCATCG 435

Db 134 GluGlyLeuProAsnGlnHisLeuGluLysLysTyrPheArgGluGlyAsp---LeuGly 152

QY 436 CTAGCGGTTTTCGAGACGCTCTGTGGAAGGCTGGTCTTATGCTGTGGAGACACCTT 495

Db 153 PheGlyValAlaPheAspPheHisGlyValGlnValGlyMetCysLeuCysAsnAspArgArg 172

QY 496 CAACCGCTAACAAATACGCTTTGTATGCACAAAATGAAGAGATTTCATTGTGCGGCTTG 555

Db 173 TrpProGluValTyrArgSerLeuAlaLeuGlnGlyAlaGluLeuValLeuGlyTyr 192

QY 556 -----CCGAGCTTTAGCCTTTATCTTAATGCGCGGCAAGCCCTGGGG 597

Db 193 AsnThrProAspPheValProGlyTyrGlnGluGluProHisAlaLysMetPhe----- 210

QY 598 CCTGATGTCAATGATAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTGFACTA 657

Db 211 -----ThrHisLeuLeuSerLeuGlnAlaGlyAlaTyrGlnAsnSerValPheValAla 228

QY 658 GCGTCGTGCGCTCGTTTCACAATCCATGCATGATATGCTTTGTACAGATGACGAAAG 717

Db 229 AlaAla-----GlyLysSerGlyPhe-----GluAspGlyHis 239

QY 718 CATGCGTGTCTTCTGGCTGTGTGACACTCAGTATCATAGGGCTGATGTTGAC 777

Db 240 HisMetIle-----GlyGlySerAlaValAlaAlaProSerGlyGluIle 254

QY 778 TTGTCGCGCTCTTGCCGAAATGAAGGGTATTCCTAC--GCAACCTTGATCCT 834

Db 255 LeuAlaLysAlaAlaGlyGluGlyAspGluValValAlaValLysAlaAspIleAspMet 274

QY 835 GGAGTACGCACTCTTGCTAAATGCGCGGAGACCCCTGCTGTCATTAATCCCGTCCGAC 894

Db 275 Gly--LysProTyrLysGluSerValPheAspPheAlaHis--ArgArgProAsp 292

QY 895 ATTACTCGCTTCTAATAGATCGCAGCCCTAAA-----TTACCGGTA 936

Db 293 AlaTyrGlyIleIleAlaGluArgLysGlyArgGlyAlaProLeuProVal 309

RESULT 9

US-07-917-111-5

; Sequence 5, Application US/07917111

; Patent No. 5565344

; GENERAL INFORMATION:

; APPLICANT: Nanba, Hirokazu

; APPLICANT: Yamada, Yukio

; APPLICANT: Takano, Masayuki

; APPLICANT: Ikenaka, Yasuhiro

; APPLICANT: Takahashi, Satomi

; APPLICANT: Yajima, Kazuyoshi

; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wegner, Cantor, Mueller & Player

; STREET: 1233 20Ch Street, N.W., Suite 300

; CITY: Washington

; STATE: D.C.

; ZIP: 20036-8218

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/917,111

; FILING DATE: 19920807

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 400848/1990

; FILING DATE: 07-DEC-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 407922/1990

; FILING DATE: 27-DEC-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: JP 078840/1991

; FILING DATE: 11-APR-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/JP91/01696

; FILING DATE: 06-DEC-1991

```
ATTORNEY/AGENT INFORMATION:
; NAME: Player Esq., William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-500-23486
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-0400
; TELEFAX: 202-835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-917-111-5
```

## Alignment Scores:

Pred. No.:	1.86e-06	Length:	311
Score:	139.00	Matches:	80
Percent Similarity:	38.3%	Conservative:	49
Best Local Similarity:	23.7%	Mismatches:	146
Query Match:	7.6%	Indels:	62
DB:	1	Gaps:	15

US-09-751-299-3 (1-1014) x US-07-917-111-5 (1-311)

```
QY 13 ATCAAGGTCGCGCTGCGAAGCCCGCCGATCTACATGGAT-----TTGAGCGCAGC 66
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 4 ValAsnAlaAlaAlaAlaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23

QY 67 GTGACAAACCATTGATGATGGAAGACGACGACGTAATATGCTCGTCGATCGCC 126
   ||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 24 ValArgArgLeuIleAlaLeuMetArgGluAlaLysAlaArgGlySerAspLeuValVal 43

QY 127 TTTCGGAAACTTGGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCAGCATGG 186
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 44 PheThrGluLeuAlaLeuThrPhe-----PheProArgTrp 56

QY 187 GCAATGCAATTTGTACGCCAATACCATGAGAAGCTCATTTGAGTTGGAT----- 234
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrGlu 69

QY 235 -----GGCCCTCAAGCTAAGCGCATTTCAATGCAGCCAGCGGTGGGAATC 282
   |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 70 LysGluMetProGlyProGluThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89

QY 283 ATGGTCACCTGGGGATGAGTAA-----CGGGTCGGTGGC-----ACCCTTTACATC 330
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 90 GlyPheTyrLeuGlyTyrAlaGluLeuAlaGluGlyGlyArgLysArgArgPheAsn 109

QY 331 AGTCAGTGTTTCATAGCGCATATGCTGACACACCATTTGGGGCCCGGAAAGTTG----- 384
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrArgLysValHisLeu 129

QY 385 -----AAACCTACTTTTGTGAACGTAATTGTTTC 414
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 ProGlyHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrPhe 149

QY 415 GGCGAAGGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGGAAGCGGTGGC 474
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 150 GluProGlyAsp--LeuGlyPheGlyValTrpArgAlaPheAspGlyValMetGlyMet 168

QY 475 TTATGCTGTGTGGAGACACCTTCAACCGCTAACAATAACGCTTGTATGCACAAATGAA 534
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 169 CysIleCysAsnAspArgArgTrpProGluThrTyrArgValMetGlyLeuGlnGlyVal 188

QY 535 GAGATTCATTGTGCGGCTTG-----CCGAGCTTTAGCCTTTATCCTAATCGCGGAAA 588
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 189 GluMetValMetLeuGlyTyrAsnThrProTyrAspHisThrGlyHisAspAspIleAsp 208

QY 589 GCCCTGGGGCCTGATGTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGC 648
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrGlnAsnSerThr 228
```

```
QY 649 TTCGTAAGCGCTGCTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTGTACAGAT 708
   ::|||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 229 TrpValIleGlyThrAlaLys-----CysGlyThr 238

QY 709 GACGAAAAGCATGCGCTGCTTCTGCGTGGTGACACTCAGTATCATAGGCGCTGAT 768
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253

QY 769 GGTGTGACTGTGTCGCGCTTTCGCCGAAATGAAGAGGGTATTCTCTACGCA----- 822
   |||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleIleThrAlaArgCys 273

QY 823 AACCTGATCCTGAGTACGACATCCTTGCTAAATGCGCGACAGCCCTGCTGCATTAAT 882
   ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 274 AspLeuAspMetGlyLysArg--TyrArgGluThrIlePheAspPheAlaArgHis--- 291

QY 883 TCCCGTCCGACATTACTCGCTTGCTAATAGATCGACGCCCTAAATTACCG 933
   |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 292 ArgGluProAspAlaTyrArgLeuIleValGluArgLysGlyAlaValPro 308

RESULT 10
US-07-917-111-6
; Sequence 6, Application US/07917111
; Patent No. 5565344
; GENERAL INFORMATION:
; APPLICANT: Nanda, Hirokazu
; APPLICANT: Yamada, Yukio
; APPLICANT: Takano, Masayuki
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Takahashi, Satomi
; APPLICANT: Yajima, Kazuyoshi
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO
; TITLE OF INVENTION: ACIDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wegner, Cantor, Mueller & Player
; STREET: 1233 20th Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036-8218
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/917,111
; FILING DATE: 19920807
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Player Esq., William E.
; REGISTRATION NUMBER: 31,409
; REFERENCE/DOCKET NUMBER: P-500-23486
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-0400
; TELEFAX: 202-835-0605
; TELEX: 440706
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
```

TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas  
STRAIN: KNK 003A (FERM BP-3181)  
US-07-917-111-6

Alignment Scores:  
Pred. No.: 1.86e-06 Length: 311  
Score: 139.00 Matches: 80  
Percent Similarity: 38.3% Conservative: 49  
Best Local Similarity: 23.7% Mismatches: 146  
Query Match: 7.6% Indels: 62  
DB: 1 Gaps: 15

US-09-751-299-3 (1-1014) x US-07-917-111-6 (1-311)

QY 13 ATCAAGTCGCCCTGCGTGCAGCCGCCCGATCTACATGAT-----TTGAGCGCAGC 66  
Db 4 ValAsnAlaAlaAlaAlaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23  
QY 67 GTGACAAACCATTGAGTTGATGAGAAGACGACGATTAATATGCTCTGATCGCC 126  
Db 24 ValArgArgLeuIleAlaLeuMetArgGluAlaLysAlaArgLysSerAspLeuVal 43  
QY 127 TTTCGGAAACTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCAGCATGG 186  
Db 44 PheThrGluLeuAlaLeuThrThrPhe-----PheProArgTyr 56  
QY 187 GCAATGCAATTTGTACGCCAATACATGAACTCATTTGAGTTGAT----- 234  
Db 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrGlu 69  
QY 235 -----GGCCCTCAAGCTTAAGCGCATTTCAAGATGCAGCCCAAGCGTTGGGATC 282  
Db 70 LysGluMetProGlyProGluThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89  
QY 283 ATGGTCACCTGGGGAGTGA-----CGGGTCGGTGC-----ACCCTTTACATC 330  
Db 90 GlyPheTyrLeuGlyTyrAlaGluLeuAlaGluGlyGlyArgLysArgArgPheAsn 109  
QY 331 AGTCAGTGGTTCATAGCGCATATGATGACACCATGCGGCCGCGAAAGTTG----- 384  
Db 110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrArgLysValHisLeu 129  
QY 385 -----AAACCTACTTTTGTGAACGTACTTTGTTTC 414  
Db 130 ProGlyHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrPhe 149  
QY 415 GCGGAAGGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGAAGCGTGGTGGC 474  
Db 150 GluProGlyAsp--LeuGlyPheGlyValTyrArgAlaPheAspGlyValMetGlyMet 168  
QY 475 TTATGCTGTTGGAGCACCCTTCAACCGCTACAAATATACGCTTTGTATGCACAAATGAA 534  
Db 169 CysIleCysAsnAspArgTyrProGluThrTyrArgValMetGlyLeuGlnGlyVal 188  
QY 535 GAGATTCATTGTGCGGCTTG-----CCGAGCTTTAGCCTTTATCCTAATGCGCGAAA 588  
Db 189 GluMetValMetLeuGlyTyrAsnThrProTyrAspHisThrGlyHisAspAspIleAsp 208  
QY 589 GCCCTGGGGCCTGATGTCATATAGCGGCTCTCGAATCATGCCGTTGAAGGCAATGC 648  
Db 209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrGlnAsnSerThr 228  
QY 649 TTCGTACTAGCGTGTGCGGCTGTTTCAACAATCCATGATCGATATGCTTTGTACAGAT 708  
Db 229 TrpValIleGlyThrAlaLys-----CysGlyThr 238  
QY 709 GACGAAAAGCATGCGTTGCTTCTGGCTGTGTGACACTGATATCATAGGCGCTGAT 768  
Db 768

Db 239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253  
QY 769 GGTTGACTTGTCGCGCCTCTTGCCGAAATGAGAGGGTATCTCTACGCA----- 822  
Db 254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleIleThrAlaArgCys 273  
QY 823 AACCTGATCCTGAGTAGCATCTTGCTAAATATGCGCGACACCTGCTGTCATTAT 882  
Db 274 AspLeuAspMetGlyLysArg--TyrArgGluThrIlePheAspPheAlaArgHis--- 291  
QY 883 TCCCGTCCGACATTAATGCTGCTGCTAATAGATCGACCCCTAAATTACCG 933  
Db 292 ArgGluProAspAlaTyrArgArgLeuIleValGluArgLysGlyAlaValPro 308  
RESULT 11  
US-08-479-638-5  
Sequence 5, Application US/08479638  
Patent No. 5695968  
GENERAL INFORMATION:  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Yamada, Yukio  
APPLICANT: Takano, Masayuki  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Takahashi, Satomi  
APPLICANT: Yajima, Kazuyoshi  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
TITLE OF INVENTION: ACIDS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wegner, Cantor, Mueller & Player  
STREET: 1233 20th Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20036-8218  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479, 638  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player Esq., William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P-500-23486  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-0400  
TELEFAX: 202-835-0605  
TELEX: 440706  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-479-638-5



Alignment Scores:

Pred. No.: 1.86e-06 Length: 311  
Score: 139.00 Matches: 80  
Percent Similarity: 38.3% Conservative: 49  
Best Local Similarity: 23.7% Mismatches: 146  
Query Match: 7.6% Indels: 62  
DB: 1 Gaps: 15

US-09-751-299-3 (1-1014) x US-08-479-638-5 (1-311)

QY 13 ATCAAGTCGCGCTGCGAAGCCGCCGATCTACATGGAT-----TTGAGGCGACG 66  
DB 4 ValAsnAlaAlaAlaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23  
QY 67 GTGACAAAACCATTTGATGATGAGAGACGACGTAATAATGCTTCGTCGATCGCC 126  
DB 24 ValArgArgLeuIleAlaLeuMetArgGluAlaLysAlaArgGlySerAspLeuVal 43  
QY 127 TTTCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCAGCAGCATGG 186  
DB 44 PheThrGluLeuAlaLeuThrThrPhe-----PheProArgTrp 56  
QY 187 GCAATGCAATTTGTACGCCAATACCATGAGAAGCTATTGAGTTGAT----- 234  
DB 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrglu 69  
QY 235 -----GGCCCTCAGACTAAGCGCATTTTCAGATGCACGCCAAGCGTTGGAATC 282  
DB 70 LysGluMetProGlyProGluThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89  
QY 283 ATGTCACCTGGGAGTGA-----CGGGTCGGTGGC-----ACCCTTTACATC 330  
DB 90 GlyPheTyrlleuGlyTyrlaGluLeuAlaGluGlyArgLysArgArgPheAsn 109  
QY 331 AGTCAGTGTTCATAGCGGATAATGTCACACCATTTGGGGCCGGAAGTTG----- 384  
DB 110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrlaGlyValHisLeu 129  
QY 385 -----AAACCTACTTTGTGAACGTACTTTGTTTC 414  
DB 130 ProGlyHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrlPhe 149  
QY 415 GCGCAAGGGGATGTTTCATCGCTAGCGGTTTTCAGACGTCGTGGAAGCGTGGTGGC 474  
DB 150 GluProGlyAsp--LeuGlyPheGlyValTrpArgAlaPheAspGlyValMetGlyMet 168  
QY 475 TTATGCTGTTGGAGCACCTTCAACCGCTAACAAATAACGTTTGTATGACAAATGAA 534  
DB 169 CysIleCysAsnAspArgArgTrpProGluThrTyrlaGlyValMetGlyLeuGlnGlyVal 188  
QY 535 GAGATTCATTGTGCGGCTGG-----CCGAGCTTTAGCCTTTATCCTAATGCGCGGAAA 588  
DB 189 GluMetValMetLeuGlyTyrlAsnThrProTyrlAspHisThrGlyHisAspArgIleAsp 208  
QY 589 GCCCTGGGGCTGATGTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGC 648  
DB 209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrlGlnAsnSerThr 228  
QY 649 TTCGTACTAGCGTCGTGGCGCTCGTTTACAAATCCATGATCGATATGCTTTGTACAGAT 708  
DB 229 TrpValIleGlyThrAlaLys-----CysGlyThr 238  
QY 709 GACGAAAAGCATGCGTTGCTTCTGGCTGTTGAGACACTCAGTATCATAGGCGCTGAT 768  
DB 239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253  
QY 769 GGTGGTGACTTGTGCGCGCTCTTTCGCGAAAATGAAGAGGTATTCTCTACGCA----- 822  
DB 254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleIleThrAlaArgCys 273  
QY 823 AACCTTGATCCTGAGTACGATCCTTGCTAAATGCGCGCAGACCCCTGCTGTCATTAT 882  
DB 274 AspLeuAspMetGlyLysArg--TyrlArgGluThrIlePheAspPheAlaArgHis--- 291

QY 883 TCCCGTCCGACATTACTGCTTGCTTAATAGATCGACGCCCTAAATTACCG 933  
DB 292 ArgGluProAspAlaTyrlArgLeuIleValGluArgLysGlyAlaValPro 308

RESULT 12

US-08-479-638-6  
; Sequence 6, Application US/08479638  
; Patent No. 5695968  
; GENERAL INFORMATION:  
; APPLICANT: Nanba, Hirokazu  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Yajima, Kazuyoshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
; TITLE OF INVENTION: ACIDS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/479,638  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,111  
; FILING DATE: 07-AUG-1992  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Player Esq., William E.  
; REGISTRATION NUMBER: 31,409  
; REFERENCE/DOCKET NUMBER: P-500-23486  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-887-0400  
; TELEFAX: 202-835-0605  
; TELEX: 440706  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Pseudomonas  
; STRAIN: KNK 003A (FERM BP-3181)  
; US-08-479-638-6

Alignment Scores: 1.86e-06 Length: 311  
Pred. No.: 139.00 Matches: 80  
Score: 38.3% Conservative: 49  
Percent Similarity: 23.7% Mismatches: 146  
Best Local Similarity:

Query Match: 7.6% Indels: 62  
DB: 1 Gaps: 15  
US-09-751-299-3 (1-1014) x US-08-479-638-6 (1-311)

```
QY 13 ATCAAGTCGCGCTGCGCAAGCCGCCGATCTACATGAT-----TTGAGCGGACG 66
   |||
   |||
   |||
Db 4 ValAsnAlaIaAlaIaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23
   |||
   |||
   |||
QY 67 GTGACAACCAATGAGTTGATGGAAGACGACGCAATAATAGCTGCTGTGATCGCC 126
   |||
   |||
   |||
Db 24 ValArgArgLeuIleAlaLeuMetArgGluAlaLysAlaArgLysSerAspLeuVal 43
   |||
   |||
   |||
QY 127 TTCCGGAACCTTGATTCAGGCTACCCATGTTCTTGGCTTGACTCACCAGCATGG 186
   |||
   |||
   |||
Db 44 PheThrGluLeuAlaLeuThrThrPhe-----PheProArgTyr 56
   |||
   |||
   |||
QY 187 GCAATGCAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGGAT----- 234
   |||
   |||
   |||
Db 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrGlu 69
   |||
   |||
   |||
QY 235 -----GGCCCTCAAGCTAAGCGCATTTTCAGATGCAAGCCGAGCGTTGGAATC 282
   |||
   |||
   |||
Db 70 LysGluMetProGlyProGluThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89
   |||
   |||
   |||
QY 283 ATGGTACCCCTGGGATGAGTGAA-----CGGTCGGTGGC-----ACCTTTACATC 330
   |||
   |||
   |||
Db 90 GlyPheTyrLeuGlyTyrAlaGluLeuAlaGluGluGlyArgLysArgArgPheAsn 109
   |||
   |||
   |||
QY 331 AGTCAGTGGTTCATAGCGCAATAAGTGACACACCATTTGGGCGCCGGAAGTTG----- 384
   |||
   |||
   |||
Db 110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrArgLysValHisLeu 129
   |||
   |||
   |||
QY 385 -----AACTTACTTTTGTGAACGTACTTTGTTTC 414
   |||
   |||
   |||
Db 130 ProGlyHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrPhe 149
   |||
   |||
   |||
QY 415 GCGCAAGGGATGTTTCATCGCTAGCGGTTTTCGAGACGCTCTGTGGAAGCGTGGTGGC 474
   |||
   |||
   |||
Db 150 GluProGlyAsp--LeuGlyPheGlyValTyrPargAlaPheAspGlyValMetGlyMet 168
   |||
   |||
   |||
QY 475 TTATGCTGTTGGAGACACCTTCAACCGCTAACAAATAGCCTTTGATGCACAAATGAA 534
   |||
   |||
   |||
Db 169 CysIleCysAsnAspArgTyrProGluThrTyrArgValMetGlyLeuGlnGlyVal 188
   |||
   |||
   |||
QY 535 GAGATTTCATTGTGCGGCTTG-----CCGAGCTTACGCTTATCCTAATCGCGCGCAA 588
   |||
   |||
   |||
Db 189 GluMetValMetLeuGlyTyrAsnThrProTyrAspHisThrGlyHisAspAspIleAsp 208
   |||
   |||
   |||
QY 589 GCCCTGGGGCTGATGTCATGTAAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGC 648
   |||
   |||
   |||
Db 209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrGlnAsnSerThr 228
   |||
   |||
   |||
QY 649 TTCGTACTAGCGTCGTGTCGCTGTTTACAAATCCATGATGATATGCTTTGTACAGAT 708
   |||
   |||
   |||
Db 229 TrpValIleGlyThrAlaLys-----CysGlyThr 238
   |||
   |||
   |||
QY 709 GACGAAAAGCATGCGTTGCTTCTGGCTGGTGTGACACTCAAGTATCATAGGGCGCTGAT 768
   |||
   |||
   |||
Db 239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253
   |||
   |||
   |||
QY 769 GGTGCTGACTTGTGCGCGCTCTTCCGGAATGAAGAGGTATTTCTTACGCA----- 822
   |||
   |||
   |||
Db 254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleIleThrAlaArgCys 273
   |||
   |||
   |||
QY 823 AACCTTGATCCTGGAGTACGATCCTTGTCTAAATGGCGGAGACCCCTGCTGTCTATTAT 882
   |||
   |||
   |||
Db 274 AspLeuAspMetGlyLysArg---TyrArgGluThrIlePheAspPheAlaArgHis--- 291
   |||
   |||
   |||
QY 883 TCCCGTCCCGACATTACTCGCTTCTGCTAATAGATCGACGCCCTAAATTACCG 933
   |||
   |||
   |||
Db 292 ArgGluProAspAlaTyrArgSerIleValGluArgLysGlyAlaValPro 308
   |||
   |||
   |||
```

RESULT 13  
US-08-294-871A-70  
Sequence 70, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanda, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294, 871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971, 758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917, 111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211, 641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25, 258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-294-871A-70

Alignment Scores:
Pred. No.: 1.86e-06 Length: 311
Score: 139.00 Matches: 80
Percent Similarity: 38.3% Conservative: 49
Best Local Similarity: 23.7% Mismatches: 146
Query Match: 7.6% Indels: 62
DB: 1 Gaps: 15

US-09-751-299-3 (1-1014) x US-08-294-871A-70 (1-311)

QY 13 ATCAAGTCGCTGCGTGAAGCCGCCGATCTACATGGAT-----TTGAGGGCGACG 66
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 ValAsnAlaAlaAlaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23

QY 67 GTGACAAAACCATTTGAGTTGATGGAAGACGACGTAATAATGCTGCTGATCGCC 126
   ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 ValArgArgLeuIleAlaLeuMetArgLysAlaLysAlaArgLysSerAspLeuVal 43

QY 127 TTCCGGAACCTGGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCAGCAGCATGG 186
   ||| ||| ||| ::| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 44 PheThrGluLeuAlaLeuThrThrPhe-----PheProArgTrp 56

QY 187 GCAATGCAATTGTACGCCAATACCATGAGACTCATTTGAGTTGGAT----- 234
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrGlu 69

QY 235 -----GGCCCTCAAGCTAAGCCGATTTGAGATGCAGCCCAAGCGTTGGGAATC 282
   ||| ||| ||| ::| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 LysGluMetProGlyProGluThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89

QY 283 ATGTCACCCCTGGGATGAGTGA-----CGGGTGGGTGGC-----ACCCTTTACATC 330
   ||| ||| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 90 GlyPheTyrLeuGlyTyrAlaGluLeuAlaGluGluGlyGlyArgLysArgGlyPheAsn 109

QY 331 AGTCAGTGGTTTCATAGCGGATTAATGTTGACACACCATTTGGGGCCGCGAAAGTTG----- 384
   ::| ||| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrArgLysValHisLeu 129

QY 385 -----AAACCTACTTTTGTGAACGTACTTTGTTTC 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 ProGluHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrPhe 149

QY 415 GCGGAAGGATGGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGTTGAAGGCTGGGTGCC 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 GluProGlyAsp--LeuGlyPheGlyValTrpArgAlaPheAspGlyValMetGlyMet 168

QY 475 TTATGCTGTTGGAGACACTTCAACCGCTAACAAATAACGCTTTGTATGCACAAATGAA 534
   ||| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 CysIleCysAsnAspArgArgTrpProGluThrTyrArgValMetGlyLeuGlnGlyVal 188

QY 535 GAGATTCATTTGCGCGCTTG-----CCGAGCTTTAGCCTTTATCCTAATGCGCGCAA 588
   ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GluMetValMetLeuGlyTyrAsnThrProTyrAspHisThrGlyHisAspAspIleAsp 208

QY 589 GCCCTGGGGCTGATGTAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGC 648
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrGlnAsnSerThr 228

QY 649 TTCGTACTAGCGTCGTGCGCTCGTTTCACAATCCATGATGCATATGCTTTGACAGAT 708
   ::| ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 TrpValIleGlyThrAlaLys-----CysGlyThr 238
```

```

QY 709 GACGAAAGCATGCCGTTGCTTCTGCGTGGTGGAGACTCAGTATCATAGGCTGAT 768
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253

QY 769 GGTGTGACTTGTCGCGCCTCTTCCCGAAATGAAGAGGGTATTCTTACGCA----- 822
   ||| ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleThrAlaArgCys 273

QY 823 AACCTGATCTGAGTACGATCCTTGCTAAATGCGGCGAGACCCCTGCTGCATTAT 882
   ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 274 AspLeuAspMetGlyLysArg--TyrArgGluThrIlePheAspPheAlaArgHis--- 291

QY 883 TCCCGTCCGACATTACTCGCTTGCTTAATGATCGCAGCCCTTAATTACCG 933
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 ArgGluProAspAlaTyrArgLeuIleValGluArgLysGlyAlaValPro 308

RESULT 14
US-08-876-398A-70
; Sequence 70, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NANBA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876, 398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211, 641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-876-398A-70
```

Alignment Scores:

Pred. No.:	1.86e-06	Length:	311
Score:	139.00	Matches:	80
Percent Similarity:	38.3%	Conservative:	49
Best Local Similarity:	23.7%	Mismatches:	146
Query Match:	7.6%	Indels:	62
DB:	2	Gaps:	15

US-09-751-299-3 (1-1014) x US-08-876-398A-70 (1-311)

```
QY      13 ATCAAGTCGCGCTGCGCAAGCCCGATCTACATGGAT-----TTGAGGCGACG 66
      ::      |||      |||      |||      |||      |||      |||      |||
Db      4 ValAsnAlaAlaAlaAlaGlnMetGlyProIleSerArgSerGluThrArgLysAspThr 23

QY      67 GTGGACAACCAATGAGTTGATGGAAGAAGCAGCAGTAATAATGCTCGTCTGATCGCC 126
      |||      ::      |||      |||      |||      |||      |||      |||
Db      24 ValArgArgLeuIleAlaLeuMetArgGluAlaLysAlaArgGlySerAspLeuVal 43

QY      127 TTCCCGAAACTTGATTCCAGGCTACCCATGGTTCTTGCGCTTGACTCACCAGCATGG 186
      |||      |||      |||      ::      |||      |||      |||      |||
Db      44 PheThrGluLeuAlaLeuThrPhe-----PheProArgTyr 56

QY      187 GCAATGCAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGGAT----- 234
      ::      ::      ::      ::      ::      ::      ::      ::
Db      57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrGlu 69

QY      235 -----GGCCCTCAAGCTAAGCGCATTTCAATGACGCCAAGCGTTGGATC 282
      |||      |||      |||      ::      |||      |||      |||      |||
Db      70 LysGlnMetProGlyProGlnThrGlnProLeuPheAspGluAlaLysArgLeuGluIle 89

QY      283 ATGTCACCCCTGGGATGAGTGA-----CGGTCGGTGGC-----ACCTTTACATC 330
      |||      |||      |||      ::      |||      |||      |||      |||
Db      90 GlyPheTyrLeuGlyTyrAlaGluLeuAlaGluGlyGlyArgLysArgArgPheAsn 109

QY      331 AGTCAGTGGTTCATAGGCGGATAATGTTGACACCACTTGGGGCCGCGAAAGTTG----- 384
      ::      ::      ::      ::      ::      ::      ::      ::
Db      110 ThrSerIleLeuValAspArgSerGlyArgIleValGlyLysTyrArgLysValHisLeu 129

QY      385 -----AAACCTACTTTTGTGAACGTACTTTGTTTC 414
      |||      |||      |||      ::      |||      |||      |||      |||
Db      130 ProGlnHisLysGluProGlnProGlyArgLysHisGlnHisLeuGluLysArgTyrPhe 149

QY      415 GCGCAAGGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGGAAGGCTGGCGC 474
      |||      |||      |||      ::      |||      |||      |||      |||
Db      150 GluProGlyAsp--LeuGlyPheGlyValTyrArgAlaPheAspGlyValMetGlyMet 168

QY      475 TTATGCTGTTGGAGCACCTTCAACCGCTAACAAATAACGCTTGTATGCACAATAAGAA 534
      |||      |||      |||      ::      |||      |||      |||      |||
Db      169 CysIleCysAsnAspArgArgTyrProGluThrTyrArgValMetGlyLeuGlnGlyVal 188

QY      535 GAGATTCAATTGTGGCGCTTG-----CCGAGCTTTAGCCTTTATCTTAATGCGGCGAAA 588
      |||      |||      |||      ::      |||      |||      |||      |||
Db      189 GluMetValMetLeuGlyTyrAsnThrProTyrAspHisThrGlnHisAspAspIleAsp 208

QY      589 GCCCTGGGGCTGATGTCATGTATGAGCGGCTCTCGAATCTATGCCGTTGAAGGCGAATGC 648
      ::      |||      |||      ::      |||      |||      |||      |||
Db      209 SerLeuThrGlnPheHisAsnHisLeuSerMetGlnAlaGlyAlaTyrGlnAsnSerThr 228

QY      649 TTCGTACTAGCGTCGTGTCGCTCGTTTCACAATCSAATGATCGATATGCTTTGTACAGAT 708
      ::      |||      |||      ::      |||      |||      |||      |||
Db      229 TrpValIleGlyThrAlaLys-----CysGlyThr 238

QY      709 GACGAAAGCATGCGTGTCTTGCGCTGTGTGACACTCACGTATCATAGGCGCTGAT 768
      ::      |||      |||      ::      |||      |||      |||      |||
Db      239 GluGluGlySerLysMetVal-----GlyGlnSerValIleValAlaPro--- 253

QY      769 GGTGCTGACTTGTCGCGCTCTTCCGAAATGAAGAGGTATTCTCTACCGA----- 822
      |||      ::      |||      |||      |||      |||      |||      |||
Db      254 SerGlyGluIleValAlaMetAlaCysThrIleGluAspGluIleIleThrAlaArgCys 273

QY      823 AACCTTGATCCTGGAGTAGCGATCCTTGCTAAATGGCGGACAGCCCTGCTGCTATTAT 882
      ::      |||      |||      |||      |||      |||      |||      |||
```

```
Db      274 AspleuAspMetGlyLysArg--TyrArgGluThrIlePheAspPheAlaArgHis---- 291

QY      883 TCCCGTCCGACATTAATCGCTTGCTAATAGATCGCAGCCCTAAATTACCG 933
      |||      |||      |||      |||      |||      |||      |||      |||
Db      292 ArgGluProAspAlaTyrArgLeuIleValGluArgLysGlyAlaValPro 308

RESULT 15
US-08-294-871A-46
; Sequence 46, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Nanba, Hiroyasu
; APPLICANT: Takano, Masayuki
; APPLICANT: Yajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiro
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294, 871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971, 758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917, 111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211, 641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
```



```

; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/127/AOPA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 46:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-294-871A-46

```

Alignment Scores:	
Pred. No.:	8.48e-06
Score:	133.00
Percent Similarity:	36.6%
Best Local Similarity:	22.0%
Query Match:	7.3%
DB:	1
Length:	303
Matches:	65
Conservative:	43
Mismatches:	109
Indels:	78
Gaps:	12

US-09-751-299-3 (1-1014) X US-08-294-871A-46 (1-303)

QY		58	GAGCGCAGCGGTGGACAAACCATTGAGTTGATGGAGAAGACGACGACCGTAATATATGCTCGT	117
Db		21	GlutInValValValAlaArgLeuLeuAspMetLeuThrLysAlaIaIaSerArgGlyAlaasn	40
QY		118	CTGATCCGCTTCCGGAAACTTGATTCAGGCTACGCCCTACCATGCTTCTTTGGCTTGACTCA	177
Db		41	PheIleValPheProGluLeuAlaLeuThrThrphe-----	52
QY		178	CCAGCATGGGCAATGCATTGTGTACGCCAATAACCATGAGAACTCATTTGAGTTGGAT---	234
Db		53	-----PheProArgTrpTrpPheThrAspGluAlaGluLeuAspser	66
QY		235	-----GGCCCTCAAAGCTAAGCGCATTTTCAGATGCAGCCGAACGG	273
Db		67	PheTyrgluThrGluMetProGlyProValValArgProLeuPhegluLysAlaIaGlu	86
QY		274	TTCGGAATCATGCTCACCCCTGGGGATGAGTGAACGGCTC-----GGTGGCACCC-----	321
Db		87	LeugLytleGlyPheasnLeuglYtyrAlaGluLeuValValGluGlyValIlysArg	106
QY		322	CTTTACATCAGTCAGTGGTTCATAGCGGATATGCTGACACCATTTGGGGCCCGCGGAAG	381
Db		107	ArgPheasnThrSerIleLeuValAspLysSerGlyLysIleValGlyLysTyraArgLys	126
QY		382	TTG-----AACCT--ACTTTGTGAAGT	405
Db		127	IleHisLeuProGluHisLysGluTyrgluAlaIaTyraArgProPheGlnHisLeuGluLys	146
QY		406	ACTTTGTTCCGGGAAGGGGATGTTTCATCGCTAGCGGTTTTGAGACGCTGTGTGAAG	465
Db		147	ArgTyrrPheGluProGlyAsp--LeuGlyPheProValTyraAspValAspAlaIaLys	165
QY		466	CTGGGTGGCTTATGCTGTGGAGACACCTTCAACCGCTAACAAATACGCTTGTATGCA	525
Db		166	MetGlyMetPheIleCysAsnAspArgArgTrpProGluAlaITrapArgValMetGlyLeu	185
QY		526	CADAATGAAGATTCATTGTGCGGCTTG-----CCGAGCTTT---	564
Db		186	ArgGlyAlaGluIleIleCysGlyGlyTyraSbnThrProThrhIsAsnProSerValPro	205
QY		565	-----AGCCTTATCCTAATGCGCGGAAAAGCCCTGGGGCTGATGTC	606
Db		206	GlnHisAspHisLeuThrSerPheHisHisLeuLeuSerMetGlnAlaGlySerTyrgln	225
QY		607	AATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGTAAGCTGCTGT	666

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:38:31 ; Search time 31.9743 Seconds  
(without alignments)  
2650.121 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1826  
Sequence: 1 atgaaagaagctatcaaggt.....cgcaactcgaagaatttga 1014

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/abs/ABSSWB spool/US09751299/runat\_26042006\_090042\_18780/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_Main -QMT=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US09751299 @CGN\_1\_1\_405@runat\_26042006\_090042\_18780 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_AA\_Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765	96.7	337	3	US-09-751-299-4	Sequence 4, Appli
2	1761	96.4	337	4	US-10-146-772-386	Sequence 386, App
3	1761	96.4	337	4	US-10-241-742-386	Sequence 386, App
4	1761	96.4	337	4	US-10-440-523-386	Sequence 386, App
5	1761	96.4	337	4	US-10-440-503-386	Sequence 386, App
6	1761	96.4	337	4	US-10-461-925-386	Sequence 386, App
7	1339	73.3	337	4	US-10-146-772-48	Sequence 48, Appl
8	1339	73.3	337	4	US-10-146-772-58	Sequence 58, Appl
9	1339	73.3	337	4	US-10-241-742-48	Sequence 48, Appl
10	1339	73.3	337	4	US-10-241-742-58	Sequence 58, Appl
11	1339	73.3	337	4	US-10-440-523-48	Sequence 48, Appl

12	1339	73.3	337	4	US-10-440-523-58	Sequence 58, Appl
13	1339	73.3	337	4	US-10-440-503-48	Sequence 48, Appl
14	1339	73.3	337	4	US-10-440-503-58	Sequence 58, Appl
15	1339	73.3	337	4	US-10-461-925-48	Sequence 48, Appl
16	1339	73.3	337	4	US-10-461-925-58	Sequence 58, Appl
17	1139	62.4	333	4	US-10-146-772-292	Sequence 292, App
18	1139	62.4	333	4	US-10-241-742-292	Sequence 292, App
19	1139	62.4	333	4	US-10-440-523-292	Sequence 292, App
20	1139	62.4	333	4	US-10-440-503-292	Sequence 292, App
21	1139	62.4	333	4	US-10-461-925-292	Sequence 292, App
22	1136	62.2	332	4	US-10-146-772-140	Sequence 140, App
23	1136	62.2	332	4	US-10-241-742-140	Sequence 140, App
24	1136	62.2	332	4	US-10-440-523-140	Sequence 140, App
25	1136	62.2	332	4	US-10-440-503-140	Sequence 140, App
26	1136	62.2	332	4	US-10-461-925-140	Sequence 140, App
27	1091	59.7	355	4	US-10-146-772-306	Sequence 306, App
28	1091	59.7	355	4	US-10-241-742-306	Sequence 306, App
29	1091	59.7	355	4	US-10-440-523-306	Sequence 306, App
30	1091	59.7	355	4	US-10-440-503-306	Sequence 306, App
31	1091	59.7	355	4	US-10-461-925-306	Sequence 306, App
32	1081	59.2	345	4	US-10-146-772-334	Sequence 334, App
33	1081	59.2	345	4	US-10-241-742-334	Sequence 334, App
34	1081	59.2	345	4	US-10-440-523-334	Sequence 334, App
35	1081	59.2	345	4	US-10-440-503-334	Sequence 334, App
36	1081	59.2	345	4	US-10-461-925-334	Sequence 334, App
37	1051	57.6	334	4	US-10-146-772-6	Sequence 6, Appli
38	1051	57.6	334	4	US-10-241-742-6	Sequence 6, Appli
39	1051	57.6	334	4	US-10-440-523-6	Sequence 6, Appli
40	1051	57.6	334	4	US-10-440-503-6	Sequence 6, Appli
41	1051	57.6	334	4	US-10-461-925-6	Sequence 6, Appli
42	1042	57.1	353	4	US-10-146-772-212	Sequence 212, App
43	1042	57.1	353	4	US-10-241-742-212	Sequence 212, App
44	1042	57.1	353	4	US-10-440-523-212	Sequence 212, App
45	1042	57.1	353	4	US-10-440-503-212	Sequence 212, App

ALIGNMENTS

RESULT 1  
US-09-751-299-4  
; Sequence 4, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; APPLICANT: Chaplin, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; PRIOR APPLICATION NUMBER: 2000-12-29  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
; OTHER INFORMATION: environmental sample  
US-09-751-299-4

Alignment Scores:  
Pred. No.: 2.03e-177  
Score: 1765.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.7%  
DB: 3  
Length: 337  
Matches: 337  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0

US-09-751-299-3 (1-1014) x US-09-751-299-4 (1-337)

QY 1 ATGAAAGAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGGATTGGAG 60  
Db 1 MetlysgluAlaIlelysvAlaIacysValGlnAlaIaProIleTyrMetAspLeuGlu 20  
QY 61 GCGACGGTGGACAAAACCATTTGAGTTGATGGAAGAAGCAGCACGCTAATATGCTGCTGTG 120  
Db 21 AlaThrValAspIysThrIleGluLeuMetCgluAlaIaAArgAsnAsnAlaArgLeu 40  
QY 121 ATCGCCTTCCGAAACTTGGATTCCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
Db 61 AlaTrpAlaMetCgluPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCCAAGCCGTTGGGAATCATGTGTCACCTGGGGATG 300  
Db 81 GlnAlaIysArgIleSerAspAlaAlaIalysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGGACACCTTTACATCAGTCAGTGGTTCATAGGCCGATAATGCTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCCGGCGAAAGTTGAACCTTACTTTTGTGAACGTACTTTGTTGGCGGAA 420  
Db 121 ThrIleGlyAlaArgArgIysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGCTCTGTGGAAGGCTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyLeuCys 160  
QY 481 TGTGGAGACACCTTCAACCGCTAACAAAATACGCTTGTATGCACAAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATTGTGCGGCTTGGCGCAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaIalysAlaLeuGlyPro 200  
QY 601 GATGTCAATGATGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTGTACTAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240  
QY 721 GCGTGTCTTCTGGCTGGTGGTGGACACTCACGTATCATAGGGCTGATGGTGGTGAATTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTCCGCGCTTGGCCGAAAATGAAGAGGGTATTCTTACGCCAAACCTTGATCCTGGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGTAAAAATGGCGGAGAGACCCGTGCTGTCATTAATCCCGTCCCGCATTAAT 900  
Db 281 ArgIleLeuAlaIysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGTAAATAGATCGAGACCCCTAAATTACCGGTAGTGAATGAAGGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 2  
US-10-146-772-386

; Sequence 386, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desautels, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-386  
  
Alignment Scores:  
Pred. No.: 5,4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 4 Gaps: 0  
  
US-09-751-299-3 (1-1014) x US-10-146-772-386 (1-337)  
QY 1 ATGAAAGAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGGATTGGAG 60  
Db 1 MetlysgluAlaIlelysvAlaIacysValGlnAlaIaProIleTyrMetAspLeuLys 20  
QY 61 GCGACGGTGGACAAAACCATTTGAGTTGATGGAAGAAGCAGCACGCTAATATGCTGCTGTG 120  
Db 21 AlaThrValAspIysThrIleGluLeuMetCgluAlaIaAArgAsnAsnAlaArgLeu 40  
QY 121 ATCGCCTTCCGAAACTTGGATTCCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
Db 61 AlaTrpAlaMetCgluPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCCAAGCCGTTGGGAATCATGTGTCACCTGGGGATG 300  
Db 81 GlnAlaIysArgIleSerAspAlaAlaIalysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGGACACCTTTACATCAGTCAGTGGTTCATAGGCCGATAATGCTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCCGGCGAAAGTTGAACCTTACTTTTGTGAACGTACTTTGTTGGCGGAA 420

Db 121 ThrTleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGTTTCATCGCTAGCGGTTTTCAGACAGCTCTGTGGAAGCGTGGGCTTATGTC 480  
Db 141 GLYAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyLeuCys 160  
QY 481 TGTGGAGACACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGlu 180  
QY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGCGCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTACTAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGCATATGCTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGlyLysHis 240  
QY 721 GCGTGTCTTGTGCTGTGTGTGACACTACGTAATAGAGCGCTGATGGTGTGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyLysPhe 260  
QY 781 GTCGCGCCTCTTGCGGAAATGAAGAGGATATTCTTACGCCAAACCTTGATCTGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGTCTAAATGGCGGACAGACCCCTGCTGTGCTATTATCCCGTCCGACATTACT 900  
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTGTCTAATAGATCGACGCCCTAAATTACCGGTAGTGAATGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337  
RESULT 3  
US-10-241-742-386  
; Sequence 386, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; PRIOR APPLICATION NUMBER: 2002-09-09  
; PRIOR FILING DATE: US/10/146,772  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29

; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRF  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-386  
Alignment Scores:  
Pred. No.: 5.4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 4 Gaps: 0  
US-09-751-299-3 (1-1014) x US-10-241-742-386 (1-337)  
QY 1 ATGAAGAAGCTATCAAGTGCCTGCGTGAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaAlaProIleTyrMetAspLeuLys 20  
QY 61 GCGACGGTGACAAACCATTTGAGTGTGATGAAGAAGACGACGATTAATGCTCGTCTG 120  
Db 21 AlathrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAlaArgLeu 40  
QY 121 ATGCGCTTTCGGGAAACTTGATTCAGGCTACCCATGTTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProThrPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGATGGATGGCCCT 240  
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCCAAGCGTTGGAAATCATGTCACCTGGGAGTG 300  
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTCCGTGGCACCCTTTACATCAGTCAGTGTGTTCAATGGGATAATGTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTCGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGTTTCATCGCTAGCGGTTTTCAGACGTCGTGTTGAAGGCTGGGCTTATGTC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyLeuCys 160  
QY 481 TGTGGAGACACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGlu 180  
QY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGCGCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTACTAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGCATATGCTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGlyLysHis 240  
QY 721 GCGTGTCTTGTGCTGTGTGTGACACTACGTAATAGAGCGCTGATGGTGTGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyLysPhe 260  
QY 781 GTCGCGCCTCTTGCGGAAATGAAGAGGATATTCTTACGCCAAACCTTGATCTGAGTA 840



|||||  
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGCTAAATGCGCGCAGACCCCTGCTGCATATATCCGTCGACATTACT 900  
Db 281 ArgIleLeuAlaIysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGCAGCCCTTAATTAACCGTAGTTGAATTTGAAGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTACGCTTTGGGTAAAGCGCTGTGAGACGGGTGCCCACTCGAAGAAATT 1011  
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 4  
US-10-440-523-386  
; Sequence 386, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/440,523  
; PRIOR FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-386

Alignment Scores:  
Pred. No.: 5.4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-440-523-386 (1-337)

QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGCAAGCCGCCGATCTACATGATTTGGAG 60  
Db 1 MetLysGluAlaIleIleYsValAlaIacysValGlnAlaAlaProIleTyrMetAspLeuLys 20  
QY 61 GCGACGGGTGACAAACCATTTGAGTTGATGGAAGAGACGACGTAATATGCTGCTG 120  
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40

QY 121 ATGCGCTTCCGAAACTTGATTCAGGCTACCATGGTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProThrPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGATGGCCCT 240  
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGAGCCAAAGCGTTGGAAATCAATGTCACCCCTGGGATG 300  
Db 81 GlnAlaIysArgIleSerAspAlaAlaIysArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTCCGTGGCACCCCTTACATCAGTCAGTGGTTCAATAGCGATAATGTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTGGCGAA 420  
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTCAATCGCTAGCGGTTTTCGAGACGCTGTGTGGAAGGCTGGGTATGC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACCTTCAACCGCTTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTATCCTAATGCGGCGAAAGCCCTGGGCGCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
QY 601 GATGTCAATGAGCGGCTCTGAAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
Db 201 AspValaAsnValAlaAlaSerArgIleTyrAlaValaGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTGTTTCACAATCCATGATTCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240  
QY 721 GCGTGTCTTCTGGCTGTGTGTGACACTACGATATGATAGGGCCCTGATGCTGACTTG 780  
Db 241 AlaLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTCGCGCCTCTTGGCCGAAATGAAAGGGTATTCTCTACGCAAACTTGATCCTGGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGCTAAATGCGCGCAGACCCCTGCTGTCATTAATCCCGTCCGACATTACT 900  
Db 281 ArgIleLeuAlaIysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGCAGCCCTTAATTAACCGTAGTTGAATTTGAAGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320

RESULT 5  
US-10-440-503-386  
; Sequence 386, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey

```
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-386

Alignment Scores:
Pred. No.: 5.4e-177 Length: 337
Score: 1761.00 Matches: 336
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.7% Mismatches: 0
Query Match: 96.4% Indels: 0
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-440-503-386 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCGCTGCGTGCAGCCGCCGATCTACATGATTTGGAG 60
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaProIleTyrMetAspLeuLys 20
QY 61 GCGACGGTGACAAAACCATTTGATTTGATGGAAGCAGCAGCTAATATGCTGCTG 120
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40
QY 121 ATGCGCTTCCGAAACTTGATTCAGGCTACAGCTAGTGTCTTCTTGCTGCTCACCA 180
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
QY 181 GCATGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGAGCCCT 240
Db 61 AlaTrpAlaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAGCGGTTGGGAATCATGTGTCACCCCTGGGATG 300
Db 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100
QY 301 AGTGAACGGGTGCGTGCAACCTTTACATCAGTCAGTGTTCATAGGCGATATGCTGAC 360
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTrpPheIleGlyAspAsnGlyAsp 120
QY 361 ACCATTGGGGCCGCGGAAAGTTGAACCTACTTTGTTGAACGTAATTTGTTGGCGAA 420
Db 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140
QY 421 GGGGATGTTTCATCCGCTAGCGGTTTTCGAGACGCTCTGTTGAAGGCTGGGCTTATGC 480
Db 141 GlyAspGlySerSerLeuAlaValAlaPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATT 540
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyrAlaLeuTyrAlaGlnAsnGluIle 180
QY 541 CATGTGCGGCTTGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCT 600
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200
QY 601 GATGTCAATGAGCGGCTCTCGAATCTATGCGCTTGAAGGCAATGCTTCTACTAGCG 660
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720
```

```
|||||
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240
QY 721 GCGTTGCTTCTGGCTGTGTGTGACACTCAGCTATCATAGGGCCTGATGTGTGACTTG 780
Db 241 AlaLeuLeuLeuAlaGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260
QY 781 GTCGCGCTCTTGGCGAAATGAAGAGGGTATTTCTTACGCCAAACCTTGATCTTGAGTA 840
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280
QY 841 CGCATCCTTGTAAATGGCGGAGAGCCCTGCTGTCATTATTCCTCCGACATTACT 900
Db 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerArgProAspIleThr 300
QY 901 CGCTTGCTAATGATCGCAGCCCTAAATTAACGGTAGTTGAATTTGAAGTGATCTTCGT 960
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGlyGlyAspLeuArg 320
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATT 1011
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluIle 337

RESULT 6
US-10-461-925-386
; Sequence 386, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-386

Alignment Scores:
Pred. No.: 5.4e-177 Length: 337
Score: 1761.00 Matches: 336
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.7% Mismatches: 0
Query Match: 96.4% Indels: 0
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-461-925-386 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCGCTGCGTGCAGCCGCCGATCTACATGATTTGGAG 60
Db 1 MetLysGluAlaIleLysValAlaCysValGlnAlaProIleTyrMetAspLeuLys 20
QY 61 GCGACGGTGACAAAACCATTTGATTTGATGGAAGCAGCAGCTAATATGCTGCTG 120
Db 21 AlaThrValAspLysThrIleGluLeuMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40
QY 121 ATGCGCTTCCGAAACTTGATTCAGGCTACAGCTAGTGTCTTCTTGCTGCTCACCA 180
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
QY 181 GCATGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGAGCCCT 240
```

|||||  
Db AlatrpaIaMeGInPheValaRgInTyHISgluAsnSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACAGCCAAAGCGGTGGGAATCATGTGACCCCTGGGATG 300  
Db 81 GlnAlaIysaRgIleSerAspAlaAlaIysaRgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGACACCCCTTACATCAGTCAGTGGTTTCATAGCGCATTAATGTTGAC 360  
Db 101 SerGluArgValGlyGlyThrLeuTyrlleSerGlnTrpPheIleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAA 420  
Db 121 ThrIleGlyAlaArgaRgIysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGGTTTCATCGCTAGCGGTTTTCGAGACGTCCTGTTGGAAGCGTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACCTTTCACACCGCTAACAAATAAGCCTTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTyraIaLeuTyraIaGlnAsnGluIle 180  
QY 541 CATTTGCGGCTTGCGCCGAGCCTTTAGCCTTTATCTTAATGCGGCGAAAGCCTGGGGCCT 600  
Db 181 HisCysAlaAlaTrpProSerPheSerLeuTyProAsnAlaAlaIysAlaLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGCGCAATGCTTGTACTAGCG 660  
Db 201 AspValAsnValAlaAlaSerArgIleTyraIaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTGTTTCACATTCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGlyLysHis 240  
QY 721 GCGTGTCTTGTGCTGCTGTGACACTCACGTCATCATTAGGGCCTGATGGTGTGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyGlyAspLeu 260  
QY 781 GTCGCGCCTCTTGCCGAAATGAAGAGGGTATCTCTACGCCAAACCTTGATCTTGAGTA 840  
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuTyraIaAsnLeuAspProGlyVal 280  
QY 841 CGCATCCTTGCTTAAATGGCGGACAGCCCTGCTGCTCATTTATCCCGTCCCGACATTACT 900  
Db 281 ArgIleLeuAlaIysMetAlaAlaAspProAlaGlyHisTySerArgProAspIleThr 300  
QY 901 CGCTTGCTAATAGATCGACAGCCCTTAATTAACCGGTAGTGAATTGAAGTGAATCTTCGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320  
QY 961 CCTTAGCCTTTGGTTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATT 1011  
Db 321 ProTyraIaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337  
RESULT 7  
US-10-146-772-48  
; Sequence 48, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: NitriIases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006

; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-48  
Alignment Scores:  
Pred. No.: 2,56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1  
US-09-751-299-3 (1-1014) x US-10-146-772-48 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGCATCTACATGATTGGAG 60  
Db 1 ValLysGluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGTGGACAAACCATTTGATGATGGAAGAAGCAGCAGTAATAATGCTGCTGTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGTACGGCCATACCATGAGAACTCATTTGAGTTGGAGTGGCCT 240  
Db 61 AlaTrpGlyMetGlnPheValaArgaRgTyHISgluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACAGCCGAGCGGTGGGAATCATGGTCAACCTGGGATG 300  
Db 81 GlnAlaIysaRgIleSerGluAlaAlaGlnaRgAlaGlyIleTyraIaAlaLeuGlyTyx 100  
QY 301 AGTGAACGGGTGCGTGACACCCCTTACATCAGTCAGTGGTTTCATAGCGCATTAATGTTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyxMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAA 420  
Db 121 ThrAlaGlyLeuAlaArgaRgIysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGGTTTCATCGCTAGCGGTTTTCGAGACGTCCTGTTGGAAGCGTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACCTTTCACACCGCTAACAAATAAGCCTTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyraIaLeuTyraIaGlnAsnGluIle 180  
QY 541 CATTTGCGGCTTGCGCCGAGCCTTTAGCCTTTATCTTAATGCGGCGAAAGCCTGGGGCCT 600  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyraRgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGCGCAATGCTTGTACTAGCG 660  
:::|||||



Db 201 GluValAsnValAlaIaIaSerArgIleTyrAlaValAlGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAGAAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGCGTGTGTGTGACACTCAGCTATCATAGGCGCTGATGTTGTTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGCGGCTCTTGCCGAAATGAAGAGGATATTCTTACGCAACCTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATGCGGCGAGACCCCTGCTGTGCTATATCCCGTCCCGACATTAAT 900  
Db 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTATAGATCGAGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCCCACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 8  
US-10-146-772-58  
; Sequence 58, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitric Oxide  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 58  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-58

Alignment Scores:  
Pred. No.: 2.56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-146-772-58 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTCCGCTGCGTGAAGCCGCCCGATCTACATGATTGGAG 60  
Db 1 ValysGluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGTTGACAAACCATTTGAGTTGATGAGAAAGACAGCAGCATTAATATGCTGCTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGlnAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACAGGCTATGGTTCTTGTGCTTGACTCACA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTGTATCGCCCAATACCATGAGAATCATTTGAGTTGATGCGCCT 240  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTCAAGTGCAGCCCAAGCGGTTGGAAATCATGTGTCACCTGGGAG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTCCGTGACACCTTTACATCAGTCAATGCTTCAATAGCGCATATGCTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTTGTAACGTACTTTGTTGGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGCGTGGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACCTTCAACCGCTAACAAATAGCTTTGTATGACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTATTCCTAATGCGGCGAAAGCCCTGGGCT 600  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTGCAATCTATGCGGTTGAAGGCAATGCTTCTGTAAGCG 660  
Db 201 GluValAsnValAlaIaIaSerArgIleTyrAlaValAlGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATGATATGCTTTGTACAGATGACGAAGAAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGCGTGTGTGTGACACTCAGCTATCATAGGCGCTGATGTTGTTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGCGGCTCTTGCCGAAATGAAGAGGATATTCTTACGCAACCTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATGCGGCGAGACCCCTGCTGTGCTATATCCCGTCCCGACATTAAT 900  
Db 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTATAGATCGAGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCCCACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 9  
US-10-241-742-48  
; Sequence 48, Application US/10241742



```

; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-241-742-48

Alignment Scores:
Pred. No.:      2,566-132      Length:      337
Score:          1339.00        Matches:      252
Percent Similarity: 84.2%      Conservative: 30
Best Local Similarity: 75.2%      Mismatches:  47
Query Match:    73.3%         Indels:       6
DB:             4             Gaps:         1

US-09-751-299-3 (1-1014) x US-10-241-742-48 (1-337)

QY      1 ATGAAGAAGCTATCAAGGTCGCTGCGTGAAGCGCCCGATCTACATGATTGGAG 60
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB      1 VallysGLuAlaIlelySValAlaCysValGlnAlaIaProValPheLeuAspLeuAsp 20

QY      61 GCGACGGTGACAAACCATTTGATTGATTGAGAGAGACAGCAGTAAATGCTCGTCTG 120
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      21 AlAThrValAspLyThrValAlaIaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40

QY      121 ATCGCCTTCCGAAACTTGATTCCAGGCTACCGCATGTTCTTGGCTTGACTCAACA 180
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      41 IleAlaPheProGluThrTriPleProGlyTyrProTriPheLeuTriPleuAspSerPro 60

QY      181 GCATGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      61 AlaTriPglYmetGlnPheValAlaArgTyrHisGluAsnSerLeuValLeuAspSerPro 80

QY      241 CAAGCTAAGCGCATTTGAGATGACGCCAAGCGGTGGGAATCATGTGACCCCTGGGGATG 300
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      81 GlnAlaIysArgIleSerGluAlaAlaIaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100

QY      301 AGTGAACGGGTGCGTGACCCCTTTACATCAGTCAGTGCTTCATAGCGGATAAGTGAGAC 360
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTriPleuIleAspAspLySgIyGlu 120

QY      361 ACCATTGGGGCGCGGCAAAAGTTGAACCTACTTTGTTGAACGTACTTTGTCGGCGAA 420
```

```

DB      121 ThrAlaGlyLeuArgArgLySLeuLyProThrHisValGluArgThrLeuPheGlyGlu 140
QY      421 GGGGATGTTTCATCGCTAGCGGTTTTCAGACAGCTCTGTTGGAAGGCTGGGCTTATGCG 480
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLeuCys 160

QY      481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGACAAATGAAGAGATT 540
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      161 CysTriPgluHisLeuGlnProLeuSerLySTyrAlaLeuTyrAlaGlnAsnGluGluIle 180

QY      541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGCAAGCCCTGGGCCT 600
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      181 HisPheAlaAlaTriPProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200

QY      601 GATGTCAATGTAGCGGCGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTGCTACTAGCG 660
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB      201 GluValAsnValAlaAlaSerArgIleTyrAlaValAlaGluGlyGlnCysPheValLeuAla 220

QY      661 TCGTGTGCGCTGTTTCAACAATCCATGATGATATGCTTGTGTACAGATGACGAAAGCAT 720
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGlySerLySHis 240

QY      721 GCGTTGCTTCTGCTGCTGTGTGTGAGACTCAGTATCATATAGGGCCTGATGCTGACTTG 780
      ::::::::::|:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260

QY      781 GTCGCGCCTTGTCCGAAATGAAGAGGGTATTCTCTACGCAAACTTGATCCTGGAGTA 840
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280

QY      841 CGCATCCTTGTGTAATGCGGCGAGACCCGTGCTGTCATTATCCCGTCCGACATTACT 900
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      281 ArgIleTyrAlaIysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300

QY      901 CGCTTGCTAATGATCGCAGCCCTTAATTAACCGGTAGTTGAAATTGAAGGTGATCTTCGT 960
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319

QY      961 CCTTACGCTTTGGGTAAAGCGTCTGAGAGCGGCTGCCAACTCGAA 1005
      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB      320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 10
US-10-241-742-58
; Sequence 58, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
```

; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-58

Alignment Scores:
Pred. No.: 2.56e-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-241-742-58 (1-337)

QY 1 ATGAAGAAGCTATCAAGTCGCGCTGCGTCAAGCCGCCGATCTACATGATTGGAG 60
:::|||||
DB 1 ValysglualailelysvalaIaCysvalGlnAlaIaIaProvalPheLeuAspLeuAsp 20

QY 61 GCGACGGTGGACAAACCATTGAGTTGATGGAAGAGCAGACGTAATATGCTGCTG 120
|||||
DB 21 AlathrValAspLysThrValAlaLeuilegluGlnAlaIaArgAsnGlyAlaArgLeu 40

QY 121 ATCGCCTTCCGAAACTTGGAATCCAGGCTACCCATGGTTCTTGGCTTGACTGACCA 180
|||||
DB 41 IleAlaPheProgluThrTrpIleProglYtyrProTrpPheLeuTrpLeuAspSerPro 60

QY 181 GCATGGGCAATGCATTTGTACGGCAATACCATGAGAACTCATGGAGTTGGATGGCCCT 240
|||||
DB 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGlnAsnSerLeuValLeuAspSerPro 80

QY 241 CAAGCTAAGCGCATTTCAGATGCAGCCAAAGCGGTGGGAATCATGCTACCCCTGGGATG 300
|||||
DB 81 GlnAlaLysArgIleSerGlnAlaIaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100

QY 301 AGTGAACGGGTCGGTGGCACCCCTTTATCATCAGTCAGTGGTTCATAGCGGATAATGGTAC 360
|||||
DB 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120

QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTAACGTACTTTGTTGGCGGAA 420
|||||
DB 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGlnArgThrLeuPheGlyGlu 140

QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGGCTTATGC 480
|||||
DB 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLeuCys 160

QY 481 TGTGGGAGACACCTTCAACCCGCTTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540
|||||
DB 161 CysTrpGlnHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGlnGluIle 180

QY 541 CATGTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTTAATGGCGGAAAGCCCTGGGCGCT 600
|||||
DB 181 HisPheAlaIaIaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200

QY 601 GATGTCATATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTACTGCGG 660
:::|||||
DB 201 GluValAsnValAlaIaIaSerArgIleTyrAlaValaGluGlyGlnCysPheValLeuAla 220

QY 661 TCGTGTGGCTCGTTTCACAATCCATGATGCATATGCTTTGTACAAGTGAAGAAAGCAT 720
|||||
DB 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240

QY 721 GCGTTCCTTCTGGCTGGTGTGACACTCAGCTATCATAGGCGCTGATGGTGGTACTTG 780
:::|||||
DB 241 SerLeuLeuGlnAlaIaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260

QY 781 GTGCGCCTCTTGCCGAAATGAAGAGGGTATTCTCTACGCCAAACCTTGATCTGGAGTA 840
|||||
DB 261 AlaArgProLeuGlyGlnAsnGlnGluGlyIleLeuTyrAlaThrLeuAspProAlaIa 280

QY 841 CGCATCCTTGCTAAATAGCGCGAGACCCCTGCTGTCATTAATCCCGTCCGACATTACT 900
|||||
DB 281 ArgIleTyrAlaIaIaIaIaIaAspProAlaGlnHisTyrSerArgProAspValThr 300

QY 901 CGCTTGCTAATAGATCGACGCCCTAATTAATACCGGTTAGTTGAATTGAAGTGATCTTCGT 960
|||||
DB 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319

QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACCGGTGGCAACTCGAA 1005
|||||
DB 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 11
US-10-440-523-48
; Sequence 48, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nicotilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-48

Alignment Scores:
Pred. No.: 2.56e-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-440-523-48 (1-337)

QY 1 ATGAAGAAGCTATCAAGTCGCGCTGCGTCAAGCCGCCGATCTACATGATTGGAG 60
:::|||||
DB 1 ValysglualailelysvalaIaCysvalGlnAlaIaIaProvalPheLeuAspLeuAsp 20

QY 61 GCGACGGTGGACAAACCATTGAGTTGATGGAAGAGCAGACGTAATTAATGCTGCTG 120
|||||

Db 21 AlaThrValAspIysThrValAlaLeuIleGluValAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATCGCCTTCCGGAACCTTGATTCACAGGCTACCCATGGTTTCTTGGCTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTTGGAGTTGGAGCCCT 240  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAAAGCGGTGGGAATCATGTGTCACCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAGTGGTTTCATAGCGCATATGCTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTGAACCTTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGAAGGCTGGGTGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysCys 160  
QY 481 TGTGGAGACACCTTCAACCGCTTAACAAATACGCTTGTATGCACAAATGAAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluIle 180  
QY 541 CATTTGGCGCTTGCCGAGCTTTAGCCTTATCCCTAATGCGCGAAAGCCCTGGGCGCT 600  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTTCGAATCTATGCCGTTGAAGGCAATGCTTCGTAACGC 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTCCAAATCCATGCATGCATATGCTTGTACAGATGACGAAACAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGCGTGGTGGTGGACACTCACGATCATAGGCGCTGATGGTGGTACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTCCGCGCTCTTGCCGAAATGAAGAGGATATTCTTACGCAACCTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGTCTAAATGGCGGAGACCCCTGCTGCTCATTAATCCCGTCCGACATTAAT 900  
Db 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

APPLICANT: Madden, Mark  
APPLICANT: Burk, Mark  
TITLE OF INVENTION: NitriLases  
FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
CURRENT APPLICATION NUMBER: US/10/440,523  
CURRENT FILING DATE: 2003-05-15  
PRIOR APPLICATION NUMBER: US/10/146,772  
PRIOR FILING DATE: 2002-05-15  
PRIOR APPLICATION NUMBER: US 60/309,006  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US 60/351,336  
PRIOR FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: US 60/300,189  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 09/751,299  
PRIOR FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 58  
LENGTH: 337  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-58  
Alignment Scores:  
Pred. No.: 2,56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1  
US-09-751-299-3 (1-1014) x US-10-440-523-58 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTCCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 ValLysGluAlaIleLeysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGTGGACAAACCATTTGAGTTGATGAGAAAGACGACACGTAATGATGCTGCTG 120  
Db 21 AlaThrValAspLysThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATCGCCTTCCGGAACCTTGATTCACAGGCTACCCATGGTTTCTTGGCTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGCAATGCATTTGTATGCGCAATACCATGAGAACTCATTTGGAGTTGGAGCCCT 240  
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCAAAGCGGTGGGAATCATGTGTCACCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAGTGGTTTCATAGCGCATATGCTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGCGAAAGTGAACCTTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGAAGGCTGGGTGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysCys 160  
QY 481 TGTGGAGACACCTTCAACCGCTTAACAAATACGCTTGTATGCACAAATGAAGATT 540

Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluLe 180  
QY 541 CATGTGCGGCTTGCCGAGCTTTAGCCTTATCCTAATGCGCGGAAAGCCCTGGGGCT 600  
Db 181 HisPheAlaIaIaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTACTAGCG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValaGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACGTATCATAGGCGCTGATGGTGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGCGGCTCTTGCCGAAATGAAGGGTATTCTCTACGCAAACTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATAGCGGAGACCTGCTGTGTCATTTATCCGCTCCGACATTAAT 900  
Db 281 ArgIleTyrAlaIaIaIaThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTAACCGGTAGTTGAATTTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValaGluValaGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValaGlu 329

RESULT 13  
US-10-440-503-48  
; Sequence 48, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; FILE REFERENCE: NITRILES AND/OR CYANIDE  
; CURRENT APPLICATION NUMBER: US/10-440,503  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; PRIOR FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-48

Alignment Scores:  
Pred. No.: 2.56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-440-503-48 (1-337)  
QY 1 ATGAAGAAGCTATCAAGTTCGCTGCGTGAAGCCGCCGATCTACATGATTTGAG 60  
Db 1 ValYsGluAlaIaIeLysValAlaCysValaGlnAlaAlaProValPheLeuAspLeuAsp 20  
QY 61 GCGACGGTGCACAAACCATTTGAGTTGATGGAAGACGACACGTAATAATGCTCGCTG 120  
Db 21 AlaThrValaAspLysThrValaAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATGCGCTTTCGGAACCTTGATTTCCAGGCTACCCATGCTTCTTTGGCTTGACTACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60  
QY 181 GCATGGCAATGCAATTTGTACGCCAATACCATAGAACTCATTTGAGTTGATGGCCT 240  
Db 61 AlaTrpGlyMetGlnPheValaArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAGCCCAAGCGGTTGGAAATCATGCTCACCCCTGGGATG 300  
Db 81 GlnAlaLysArgIleSerGluAlaAlaGlnaArgAlaGlyIleTyrValaAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGCGTGACCCCTTACATCAGTCAGTGTTTCATAGCGATAAATGGTGAC 360  
Db 101 SerGluArgValaSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
QY 361 ACCATTGGGGCCCGGCGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTTCCGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValaGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGCTGTGGAAGGCTGGGTGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGAGACACCTTGAACCGCTAACAATAACGCTTGTATGCACAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180  
QY 541 CATGTGCGGCTTGCCGAGCTTTAGCCTTATCCTAATGCGCGGAAAGCCCTGGGGCT 600  
Db 181 HisPheAlaIaIaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGTAACG 660  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValaGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240  
QY 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACGTATCATAGGCGCTGATGGTGTGACTTG 780  
Db 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTGCGGCTCTTGCCGAAATGAAGGGTATTCTCTACGCAAACTTGATCCTGGAGTA 840  
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 CGCATCCTTGCTAAATAGCGGAGACCTGCTGTGTCATTTATCCGCTCCGACATTAAT 900  
Db 281 ArgIleTyrAlaIaIaIaThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTAACCGGTAGTTGAATTTGAAGTGATCTTGGT 960  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValaGluValaGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValaGlu 329

RESULT 14  
US-10-440-503-58  
; Sequence 58, Application US/10440503



Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE  
; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440,503  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; PRIOR FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 58  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-58

Alignment Scores:  
pred. No.: 2.56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-440-503-58 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCGCTGCGTGAAGCGCCCGATCTACATGATTGGAG 60  
:::|||||  
Db 1 VallysgluAlaIlelyValAlaCysValGlnAlaAlaProValPheLeuAspLeu 20  
QY 61 GCGACGTGACAAACCATTTGATGTGAGAGACACGTAATAATGCTGCTG 120  
|||||  
Db 21 AlaThrValAspIleThrValAlaLeuIleGluGluAlaAlaArgAsnGlyAlaArgLeu 40  
QY 121 ATCGCCTTCCGAAACTTGATTCAGAGCTACCCATGCTTCTTGCTTGACTCACA 180  
|||||  
Db 41 IleAlaPheProGluThrIleProGlyTyrProIlePheLeuThrLeuAspSerPro 60  
QY 181 GCATGGGCAATGCAATTTGACGCCAATACCATGAGAACTCATTTGAGTTGGAGCCCT 240  
|||||  
Db 61 AlaTrpGlyMetGlnPheValAlaArgTyrHisGluAsnSerLeuValLeuAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTCAAGTCAGCCCAAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
|||||  
Db 81 GlnAlaIleArgIleSerGluAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100  
QY 301 AGTGAACGGGTGCGTGGACCCCTTACATCAGTCAGTGTTCATAGCGCATTAATGTTGAC 360  
|||||  
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspIleGlyLeu 120  
QY 361 ACCATTGGGGCCCGGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTTGGCGAA 420  
|||||  
Db 121 ThrAlaGlyLeuAlaArgIleGlyLeuIleProThrHisValGluArgThrLeuPheGlyGlu 140  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTTGGAAGGCTGGGTGCTTATGC 480  
|||||  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
QY 481 TGTGGGAGACCTTCAACCGCTAAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
|||||  
Db 161 CysTrpGluHisLeuGlnProLeuSerIleTyrAlaLeuTyrAlaGlnAsnGluGluIle 180

QY 541 CATGTGCGCTTGCCGAGCTTTAGCCTTATCCTAATGCGCGAAAGCCCTGGGCGCT 600  
|||||  
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
:::|||||  
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValAlaGluGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGCGCTGCTTTCACATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
|||||  
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerIleHis 240  
QY 721 GCGTGTCTTGTGCTGCTGCTGTGAGACACTCAGTATCATAGGCGCTGATGCTGACTTG 780  
:::|||||  
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260  
QY 781 GTCGCGCTTGTGCGGAAATGAAGAGGATTTCTTACGCAAACTTGATCCTGAGTA 840  
|||||  
Db 261 AlaArgProLeuGlyGluAsnGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280  
QY 841 GCGATCCTTGTAAATGCGGAGACCTGCTGCTCATTTATCCCGTCCGACATTACT 900  
|||||  
Db 281 ArgIleTyrAlaIleThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300  
QY 901 CGCTGTAAATAGATCGAGCCCTAAATTAACCGTAGTTGAATTTGAAGGTGATCTTCGT 960  
|||||  
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319  
QY 961 CCTTACGCTTTGGGTAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

RESULT 15  
US-10-461-925-48  
; Sequence 48, Application US/10461925  
; Publication No. US20040053378A1

; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001  
; CURRENT APPLICATION NUMBER: US/10/461,925  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,317  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/392,944  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-48

Alignment Scores:  
pred. No.: 2.56e-132 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-461-925-48 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCGCTGCGTGAAGCGCCCGATCTACATGATTGGAG 60  
:::|||||  
Db 1 VallysgluAlaIlelyValAlaCysValGlnAlaAlaProValPheLeuAspLeu 20

```

QY 61 GCGACGGTGACAAACCATTTAGTTGATGGAAGACGACGTAATATGCTGCTG 120
    |||||
Db 21 AlaThrValAspLysThrValAlaLeuIleGluAlaAlaArgAsnGlyAlaArgLeu 40
    |||||
QY 121 ATGCGCTTTCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 180
    |||||
Db 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
    |||||
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCT 240
    |||||
Db 61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80
    |||||
QY 241 CAAGCTAAGCGCATTTACAGATGACCCAAAGCGGTTGGAAATCATGGTCACCCCTGGGGATG 300
    |||||
Db 81 GlnAlaLysArgIleSerGlnAlaIaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100
    |||||
QY 301 AGTGAACGGGTCGGTGCGACCCCTTACATCAGTCAGTGGTTGTAACGTACTTTGTCGGCGAA 360
    |||||
Db 101 SerGluArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120
    |||||
QY 361 ACCATTGGGGCCCGCGGAAGTTGAAACCTACTTTTGTGAACGTACTTTGTCGGCGAA 420
    |||||
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140
    |||||
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTTGGAAGCGCTGGTGCTTATGC 480
    |||||
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLysLeuCys 160
    |||||
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATATACGCTTTGTATGCACAAATGAAGAGATT 540
    |||||
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluGluIle 180
    |||||
QY 541 CATTTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAAATGCGCGAAAGCCCTGGGGCCT 600
    |||||
Db 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200
    |||||
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTACTAGCG 660
    |||||
Db 201 GluValAsnValAlaAlaSerArgIleTyrAlaValAlaGluGlyGlnCysPheValLeuAla 220
    |||||
QY 661 TCGTGTGGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720
    |||||
Db 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240
    |||||
QY 721 GCGTTGCTTCTGGCTGGTGGTGGAACACTACGATATCATAGGCGCTGATGGTGCTGACTTG 780
    |||||
Db 241 SerLeuLeuGlnAlaGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260
    |||||
QY 781 GTGCGGCTCTTGCCGAAATGAAGAGGATTTCTTACGCAAAACCTTGATCCTGGAGTA 840
    |||||
Db 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280
    |||||
QY 841 CGCATCCTTGCTAAATGCGGCGAGACCCCTGCTGTCATTATCCCGTCCGACATTACT 900
    |||||
Db 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300
    |||||
QY 901 CGCTTGCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAAATGAAGGTGATCTTCGT 960
    |||||
Db 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValAlaGlyArgGluIle--- 319
    |||||
QY 961 CTTTACGCTTTGGTTAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005
    |||||
Db 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329
    |||||

```

Search completed: April 27, 2006, 01:51:33  
 Job time : 166.872 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:41:06 ; Search time 3.94745 Seconds  
(without alignments)  
2336.425 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826  
Sequence: 1 atgaaagaagctatcaaggt.....cgcaactcgaagaatttga 1014

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 464238

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=/abs/ABSSWEB\_spool/US09751299/runat\_26042006\_090044\_18835/app\_query.fasta\_1  
-DB=Published\_Applications\_AA\_New\_QFMT=fastan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPLC=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abs02p  
-USER=US09751299@CGN\_1\_1\_47@runat\_26042006\_090044\_18835 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:  
1: /SID5/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
2: /SID5/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /SID5/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
4: /SID5/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
5: /SID5/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
6: /SID5/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
7: /SID5/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
8: /SID5/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	822	45.0	356	6	US-10-537-075-7	Sequence 7, Appli
2	780	42.7	369	6	US-10-919-182-18	Sequence 18, Appl
3	777	42.6	369	6	US-10-919-182-16	Sequence 16, Appl
4	776	42.5	369	6	US-10-919-182-4	Sequence 4, Appli
5	776	42.5	369	6	US-10-919-182-6	Sequence 6, Appli
6	774	42.4	369	6	US-10-919-182-12	Sequence 12, Appl
7	774	42.4	369	6	US-10-919-182-14	Sequence 14, Appl
8	765	41.9	369	6	US-10-919-182-8	Sequence 8, Appli
9	451	24.7	333	7	US-11-096-568A-20687	Sequence 20687, A

10	451	24.7	351	7	US-11-096-568A-20686	Sequence 20686, A
11	446.5	24.5	330	7	US-11-096-568A-6955	Sequence 6955, Ap
12	446.5	24.5	350	7	US-11-096-568A-6954	Sequence 6954, Ap
13	323	17.7	233	7	US-11-096-568A-6956	Sequence 6956, Ap
14	274.5	15.0	193	7	US-11-096-568A-20688	Sequence 20688, A
15	126	6.9	295	7	US-11-079-463-5457	Sequence 5457, Ap
16	114.5	6.3	259	7	US-11-045-004-822	Sequence 822, App
17	109	6.0	296	7	US-11-045-004-2100	Sequence 2100, Ap
18	106	5.8	460	7	US-11-087-099-8682	Sequence 8682, Ap
19	103.5	5.7	493	7	US-11-096-568A-11657	Sequence 11657, A
20	103.5	5.7	643	7	US-11-096-568A-27797	Sequence 27797, A
21	100	5.5	251	7	US-11-096-568A-18174	Sequence 18174, A
22	97	5.3	388	6	US-10-497-135-19	Sequence 19, Appl
23	97	5.3	388	7	US-11-269-215-19	Sequence 19, Appl
24	97	5.3	647	7	US-11-087-099-7887	Sequence 7887, Ap
25	97	5.3	647	7	US-11-188-298-18316	Sequence 18316, A
26	96	5.3	252	7	US-11-055-822-134	Sequence 134, App
27	96	5.3	252	7	US-11-055-822-1120	Sequence 1120, Ap
28	95	5.2	312	7	US-11-096-568A-12515	Sequence 12515, A
29	94.5	5.2	258	7	US-11-072-512-3034	Sequence 3034, Ap
30	92	5.0	153	6	US-10-986-501-173	Sequence 173, App
31	92	5.0	386	6	US-10-497-135-20	Sequence 20, Appl
32	92	5.0	386	7	US-11-269-215-20	Sequence 20, Appl
33	91.5	5.0	288	7	US-11-096-568A-16873	Sequence 16873, A
34	91.5	5.0	635	7	US-11-096-568A-29628	Sequence 29628, A
35	91.5	5.0	710	7	US-11-096-568A-29627	Sequence 29627, A
36	90.5	5.0	262	7	US-11-096-568A-17194	Sequence 17194, A
37	89.5	4.9	463	7	US-11-087-099-5333	Sequence 5333, Ap
38	88.5	4.8	380	7	US-11-096-568A-27808	Sequence 27808, A
39	88.5	4.8	388	7	US-11-096-568A-21838	Sequence 21838, A
40	88.5	4.8	413	7	US-11-096-568A-20771	Sequence 20771, A
41	88	4.8	489	7	US-11-188-298-3840	Sequence 3840, Ap
42	87.5	4.8	312	7	US-11-096-568A-9235	Sequence 9235, Ap
43	86	4.7	358	7	US-11-087-099-5821	Sequence 5821, Ap
44	86	4.7	358	7	US-11-188-298-16344	Sequence 16344, A
45	85.5	4.7	293	7	US-11-096-568A-12504	Sequence 12504, A

ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zelinski, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-10-537-075-7

Alignment Scores:  
Pred. No.: 2.43e-76 Length: 356  
Score: 822.00 Matches: 160  
Percent Similarity: 65.1% Conservative: 56  
Best Local Similarity: 48.2% Mismatches: 114  
Query Match: 45.0% Indels: 2  
DB: 6 Gaps: 1  
US-09-751-299-3 (1-1014) x US-10-537-075-7 (1-356)



QY	4	AAAGAAAGCTATCAAGGTCGGCTGCGTGAAGCCGCCCGGATCTACATGGATTGGAGGG	63
		:::::	
Db	4	ArgLysIleValArgAlaIleValGlnAlaIleSerProAsnTyrAspLeuAlaThr	23
QY	64	ACGGTGAACAACCATTTGATGGAGAAGACAGACACGTAATATGCTGCTGATC	123
		::	
Db	24	GlyValAspLysThrIleGluLeuAlaArgGlnAlaArgAspGluGlyAspLeuIle	43
QY	124	GCCTTTCCGAAACCTTGATTCAGGTAACCCATGCTTTCTTTGGCTTGAACACAGCA	183
		::	
Db	44	ValPheGluGluThrTrpLeuProGluTyrProPheHisValTrpLeuGlyAlaProAla	63
QY	184	TGGGCAATGCATTTGTACGSCCAATACAGAACTCATTTGGATGGATGGCCCTCAA	243
		::	
Db	64	TrpSerLeuLysTyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGlu	83
QY	244	GCTAAGCCGCAATTCAGATGCAGCCACGCGTTGGGAATCATGGTCAACCTGGGATGAGT	303
		:::      ::            ::	
Db	84	PheGlnArgIleAlaGlnAlaAlaArgThrLeuGluTyrPheIleAlaLeuGluTyrSer	103
QY	304	GAAACGGTCGGTGGCAACCTTACATCAGTGGTTCATAGGCGATATAGTGACACCC	363
		::	
Db	104	GluArgSerGlyGlySerLeuTyrLeuGluGlnCysLeuIleAspAspLysGlyMet	123
QY	364	ATTGGGGCCCGGCGAAAGTTGAACCTACTTTGTGAAACGTACTTTGTTCGGCGAAGG	423
		::: ::	
Db	124	LeuTrpSerArgArgLysLeuLysProThrHisValGluArgThrValPheGluGluGly	143
QY	424	GATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGGAAGGCTGGGCTTATGCTGT	483
		::    ::	
Db	144	TyrAlaArgAspLeuIleValSerAspThrGluLeuGluArgValGlyAlaLeuCysCys	163
QY	484	TGGGAGCACTTCAACCGCTAAACAATACGCTTGTATGACACAACAATAGAGAATTCAT	543
		::	
Db	164	TrpGluHisLeuSerProLeuSerLysTyrAlaLeuTyrSerGlnHisGlnAlaIleHis	183
QY	544	TGTGGCGCTTGGCCGAGCTTAGCCTTTATCCTAATGCGGCGAAGCCCTGGGCGCTGAT	603
Db	184	IleAlaAlaTrpProSerPheSerLeuTyrSerGluGlnAlaHisAlaLeuSerAlaLys	203
QY	604	GTCATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTAAGCTCG	663
		::            ::	
Db	204	ValAsnMetAlaAlaSerGlnIleTyrSerValGluGlnCysPheThrIleAlaAla	223
QY	664	TGTGGCGCTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAGAAGCATGCG	723
		:::::       ::	
Db	224	SerSerValValThrGlnGluThrLeuAspMetLeuGluValGlyLuhHisAspAlaPro	243
QY	724	TTCGTTCTGGCTGGTGGTGGACATCAGTATCATTAGGGCTGATGGTGGTGAATGGTC	783
Db	244	LeuLeuLysValGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAla	263
QY	784	GCGGCTCTGCGCAAAATGAGAAGGATATCTGTACGCAAACTTGATCCTGAGTACGC	843
		::            ::	
Db	264	ProTyrLeuProHisAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluIle	283
QY	844	ATCCTTGCSTAATAATGGCGGACAGACCTGCTGTCATTAATCCCGTCCCGACATTAATCCG	903
		::	
Db	284	AlaPheAlaLysAlaIleAsnAspProValGluHisTyrSerLysProGluAlaThrArg	303
QY	904	TTCGTAATAGATCGCAGCCCTAATTAACCGGTAGTTGAATTTGAAGGTGATCTTCGCTCT	963
		::          ::	
Db	304	LeuValLeuAspLeuGluHisArgAspPrometThrArgValHisSer-----LysSer	321
QY	964	TACGCTTTGGGTAAGCGCTGTAGACGGGTGCGCA	999
		:::	
Db	322	ValThrArgGluGluAlaProGluGlnGlyValGln	333

RESULT 2  
US-10-919-182-18  
; Sequence 18, Application US/10919182  
; Publication No. US20060035352A1

```

; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change
US-10-919-182-18

```

US-09-751-299-3 (1-1014) x US-10-919-182-18 (1-369)

QY 22 GCCTGCGTCAAGCCGCCCGATCTACATGTGATTGGAGCGACG6TGACAATAACCATT 81  
||| ||||||| |:::||||| ::|||::|::|::|::|  
Db 11 AlaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThrIleAspLysSerIle 30

```

QY      82 GAGTTGATGAAGAAGCAGCAGCATAATGCTGCTGATCGCTTTCGGAAACTTGG 141
          :::::|||||:::  |||  |||||:::  ::
Db      31 G|y|l|e|l|e|g|l|u|a|l|a|l|a|g|l|n|y|g|l|y|a|l|a|s|e|r|l|e|u|l|l|e|a|l|a|p|h|e|p|r|o|g|l|u|a|l|p|h|e 50

```

```
QY      142 ATTCAGGCTAACCCATGATTCTTTGGCTTGACTCACCAGCATGGCAATGC AATTGTGA 20
        |||||
Db       51 ILEProGLTYrPrOTyrrTPalatrPluenglYAspValLYSTyrSerleuSerPheThr 70
```

```

Qy      202 CGCCAAATACCATGAGAACTCATTGGAGTTGGATGGCCCTCAAGCGTAAAGCCGATTTCAGAT 26
        ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      71 SerArgIYrHtsgluAsnSerLeuGlInleuGLyAspAspArgMcetArgArgLeuGlInleu 90

```

```

Qy      262 GCAGCCCAAGCGGTTGGGAATCATGTCACCTGGGGATGAGTGAACGGGTCGGTGGCACC 32
      |||||::|||:::|||||
Db      91 A|a|a|a|a|a|g|a|n|l|y|s|i|l|e|a|l|a|e|u|a|l|m|e|t|g|l|y|r|s|e|r|g|l|u|a|g|l|s|e|r 11

```

```

Oy      322 CTTTACATCAGTCAGTGTTCATAGCGCATATGTGTGACACCACTGGGGCCCCGGCGAAG 38
      |||:::||||| ||||| ::|::: |||||
Db      111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyIuIleValAlaAsnArgArgLys 13

```

Qy 382 TTGAACCTACTTTTGTGTGAACCGTACTTTGTCGGCGAAGGGGATGGTTCAATCGTAGCG 44  
|||  
Db 131 LeuysProthrHisValGIuArgThrIeTyGLyGIuGIYasnGLyThraSPheLeu 15

```

QY      442 GTTTCGAGACGCTGTGGAGGCGTGGTGCTTATGCTGTGGAGACCTTCACCG 50
      :: :: |||||:::||||| ||||| ||||| ||||| ||||| |||||
Db      151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTyrGlnHisLeuGlnPro 177

```

```

Oy      502 CTACAAATATGCGCTTTGTATGCACAAATGAGAAGATTCTATTGCGGCTTGCGCCGAGC 56
      |||::|||::  ::|||::  |||::|||::  |||::|||::  |||::
Db      171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTyrProAla 19

```

Oy 562 TTAGC--CTTATTCCTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAGCGGC 61  
 ||| ||| |||::: ||| ::| |||  
 Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluIalaAsnAlaThrVal 20

QY           619 TCTCGAATCTATGCCGTTGAAGGGCAATGCTTGTACTAGCGTCGTGNGCCTCGTTTCA 67  
       :::| | | | | | | | | | | | | | | | | |  
Db          210 ThrArgSerTyzAlarIleGluglYgInThrPheValLeucySserThrInGlnValIleGly 22

QY 679 CAATCCATGATGATGCTTTGTACAGATGACGAAAGCATGCCGTTGCTGCTGCT 738  
DB 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluInArgAlaLeuLeuProGlnGly 249  
QY 739 GGTGACACTCAGCTATCATAGAGCCTGATGGTGTGACTTGGTCGCCCTTGCCGAA 798  
DB 250 CysGlyTrrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269  
QY 799 AATGAAGAGGCTATCTCTACGCAACCTTGATCCTGGAGTACGCATCCTTGCTAAATG 858  
DB 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuAlaLysAla 289  
QY 859 GCGGACAGACCTGCTGCTATTTATCCCGTCCCGACATTACTGCTTGCTTAATGATCGC 918  
DB 290 GlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309  
QY 919 AGCCCTAATTAACCGGTA-----GTTGAATTTGAAGGTGATCTT 957  
DB 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 3  
US-10-919-182-16  
; Sequence 16, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 16  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change  
US-10-919-182-16

Alignment Scores:  
Pred. No.: 1.05e-71 Length: 369  
Score: 777.00 Matches: 150  
Percent Similarity: 67.3% Conservative: 62  
Best Local Similarity: 47.6% Mismatches: 99  
Query Match: 42.6% Indels: 4  
DB: 6 Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-16 (1-369)

QY 22 GCCTGCGTGAAGCCGCCGATCTACATGATTGGAGCGGACGGTGACAAACCATT 81  
DB 11 AlaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGAAGAAGACGACGCAATATATGCTCGTCTGATCGCTTTCCGGAACCTTG 141  
DB 31 GlyIleIleGluGluAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCGAGGCTACCCATGCTTCTTGGCTTGACTCACCAGCATGGGCAATGCAATTGTA 201  
DB 51 IleProGlyTyrProTyrTrrPalaTrpLeuGlyAspValLysTyrSerLeuSerPheThr 70  
QY 202 CGCCAATACCATGAAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCCGATTTCAGAT 261  
DB 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCAAAGCGTTGGCAATCATGTCACCCCTGGGATGAGTGAACGGGTGGTGACCC 321  
DB 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110

QY 322 CTTTACATCATGCTGCTTATGAGGCAATATGTTGACACCATTGGGGCCCGGAAAG 381  
DB 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGAACGTAATTGTTCCGGAAGGGAAGGTTTCATCGCTAGCG 441  
DB 131 LeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTCGAGACGCTGTTGGAAGGCTGGTGCTTATGCTGTTGGAGCACTTCAACCG 501  
DB 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCystrPrluHisValGlnPro 170  
QY 502 CTACAAATAACGCTTTGATGACCAAAATGAAGAGATTCAATTGTGGGCTTGCCGAGC 561  
DB 171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTrrProAla 190  
QY 562 TTTAGC---CTTATCCTAATGCGGGAAGCCCTGGGCTGATGTCATGTAGCGGCC 618  
DB 191 MetSerProLeuGlnProAspValPheGln---LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTGAATCTATGCGGTTGAAGGCAATGCTTGACTAGCGTGTGCGCTGTTCA 678  
DB 210 ThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTCTTGCTGCTGCT 738  
DB 230 ProSerAlaIleGluThrPheCysLeuAsnAspGlyGlnArgAlaLeuLeuProGlnGly 249  
QY 739 GGTGACACTCAGCTATCATAGAGCCTGATGGTGTGACTTGGTGGGCTTGCCGAA 798  
DB 250 CysGlyTrrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269  
QY 799 AATGAAGAGGCTATCTCTACGCAACCTTGATCCTGAGTACGCAATCCTTGCTAAATG 858  
DB 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuAlaLysAla 289  
QY 859 GCGGACAGCCTGCTGCTATTTATCCCGTCCCGACATTACTCGCTTGCTAATAGATCGC 918  
DB 290 GlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309  
QY 919 AGCCCTAATTAACCGGTA-----GTTGAATTTGAAGGTGATCTT 957  
DB 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 4  
US-10-919-182-4  
; Sequence 4, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Acidovorax facilis 72W  
US-10-919-182-4

Alignment Scores:  
Pred. No.: 1.33e-71 Length: 369  
Score: 776.00 Matches: 150  
Percent Similarity: 67.0% Conservative: 61  
Best Local Similarity: 47.6% Mismatches: 100  
Query Match: 42.5% Indels: 4  
DB: 6 Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-4 (1-369)

QY 22 GCCTGCGTGCAGACCGCCCGATCTACATGGATTGGAGCGGACGGTGGACAAACCAT 81  
Db 11 AlaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGGAAGAAGCAGCACGTAATAATGCTGCTGTGATGCGCTTCCGGAACCTGG 141  
Db 31 GlyIleIleGluGluAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCAGGCTACCCATGGTTCTTGGCTTGACTCAACGACATGGGCAATGCAATTGTGA 201  
Db 51 IleProGlyTyrProTyrTrpAlaTrpLeuGlyAspValLysTyrSerLeuSerPheThr 70  
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGATGGACCCCTCAAGCTTAAGCGCATTTCA 261  
Db 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCAAGCGGTTGGGAATCATGTGCACCCCTGGGGATGATGAACGGGTGGTGACCC 321  
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGCGCATATAGTGACACCATTTGGGGCCCGGAAG 381  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGTAACGTACTTTGTTCGCGGACGAGGGGATGGTTCAATCGCTAG 441  
Db 131 LeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTTCGAGACGTCGTGGTGGAAAGGCTGGGTGCTTATGCTGTGGAGACCTTCAACCG 501  
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrpGluHisPheGlnPro 170  
QY 502 CTAACAAAATACGCTTGTGTATGCACAAATGAAGATTGTCGCGCTGGCCGAGC 561  
Db 171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTrpProAla 190  
QY 562 TTTAGC--CTTTATCCTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTCGAATCTATGCCGTGAAGGGCAATGCTTGCCTAGCGTGTGCGCTGTTTCA 678  
Db 210 ThrArgSerTyrAlaIleGluGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTTCGCTTGGCTGGT 738  
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249  
QY 739 GGTGACACTCACGTAATCATAGGGCCCTGATGGTGGTGACTTGGTCCGCTCTTGCCGAA 798  
Db 250 CysGlyTTrpAlaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269  
QY 799 AATGAAGAGGGTATCTCTACGCAAACTTGATCCTGGAGTAGCGCATCTTGCTAAATG 858  
Db 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuLeuAlaLysAla 289  
QY 859 GCGGACAGACCTGCTGTCATTATTCCTCCGTCGACATTACTGCTTGCTAATAGATCGC 918  
Db 290 GlyAlaAspProValGlnHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309  
QY 919 AGCCCTAAATTACCGTA-----GTTGAATTGAAGGTGATCTT 957  
Db 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 5  
US-10-919-182-6  
; Sequence 6, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert

; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase B2 and H9  
US-10-919-182-6

Alignment Scores:  
Pred. No.: 1.33e-71 length: 369  
Score: 776.00 Matches: 150  
Percent Similarity: 67.0% Conservative: 61  
Best Local Similarity: 47.6% Mismatches: 100  
Query Match: 42.5% Indels: 4  
DB: 6 Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-6 (1-369)

QY 22 GCCTGCGTGCAGACCGCCCGATCTACATGGATTGGAGCGGACGGTGGACAAACCAT 81  
Db 11 AlaThrValGlnAlaGluProValTrpLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGGAAGAAGCAGCACGTAATAATGCTGCTGTGATGCGCTTCCGGAACCTGG 141  
Db 31 GlyIleIleGluGluAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCAGCATGGGCAATGCAATTGTGA 201  
Db 51 IleProGlyTyrProTyrTrpAlaTrpLeuGlyAspValLysTyrSerLeuSerPheThr 70  
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGATGGACCCCTCAAGCTTAAGCGCATTTCA 261  
Db 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCAAGCGGTTGGGAATCATGTGCACCCCTGGGATGATGAACGGGTGGTGACCC 321  
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGGGCATATAGTGACACCATTTGGGGCCCGGAAG 381  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGTAACGTACTTTGTTCGCGGACGAGGGGATGGTTCAACCG 441  
Db 131 LeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTTCGAGACGTCGTGGTGGAAAGGCTGGGTGCTTATGCTGTGGAGACCACTTCAACCG 501  
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrpGluHisPheGlnPro 170  
QY 502 CTAACAAAATACGCTTGTGTATGCACAAATGAAGATTGTCGCGCTTGCCGAGC 561  
Db 171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTrpProAla 190  
QY 562 TTTAGC--CTTTATCCTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTCGAATCTATGCCGTGAAGGGCAATGCTTGCCTAGCGTGTGCGCTGTTTCA 678  
Db 210 AlaArgSerTyrAlaIleGluGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTTCGCTTGGCTGGT 738  
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249



```

QY      739  GGTGGACACTCACCGTATCATAGGGCGCTGATGGTGTGACTTGTCGGCGCTCTTGCCGAA  798
      |||  :::|||||  |||||  ::|||  |||||  |||
Db      250  CysGlyTrpAlaArgIleTyrglyProaspGlysergIleuAlaIysProLeuAlaGlu  269
      |||  :::|||||  |||||  ::|||  |||||  |||
QY      799  AATGAAGAGGTATTCTCTACGCAACCTTGATCCTGAGATAGCATCCCTGTGTAATG  858
      ::|  |||||  |||||  |||||  ::|||  |||||
Db      270  AspIaGluGlyIleLeuTyrrAlaGluIleaspHeuGluGlnIleLeuLeuAlaIysAla  289
      |||  |||||  |||||  |||||  ::|||  |||||
QY      859  GCGGAGACCCCTGCTGTCATATTATCCCGTCCCGACATTAATCTCGTTCGTAATGATCGC  918
      |||||  |||||  |||||  |||||  |||||  |||
Db      290  GlyAlaaspProValGlyHisItyrSerArgProaspValLeuSerValGlnPheaspPro  309
      |||||  |||||  |||||  |||||  |||||  |||
QY      919  AGCCCTAAATTACCGGTA-----GTTGAATTGAAGGTGATCTT  957
      |||||  |||||  ::|  |||||  |||||  |||
Db      310  ArgAsnHisThrProValHisArgIleGlyIleaspGlyArgLeu  324

```

```

RESULT 6
US-10-919-182-12
; Sequence 12, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
US-10-919-182-12

```

Alignment Scores:	
Pred. No.:	2.14e-71
Score:	774.00
Percent Similarity:	66.7%
Best Local Similarity:	47.6%
Query Match:	42.4%
DB:	6
	Gaps: 3
	Length: 369
	Matches: 150
	Conservative: 60
	Mismatches: 101
	Indels: 4

US-09-751-299-3 (1-1014) x US-10-919-182-12 (1-369)

QY 22 GCCTGCGTGAAGCCGCCCCGATTCATCATGATTGTGGAGCGGACGGTGGACAAACCATT 81  
||| ||||| ||| ::::: ||| ::||| ||| ::|||  
Db 11 AlaThrValGlnalaglProValTyrLeuAspAlaaspAlaThrIleAspYsserIle 30  
||| ::|||  
QY 82 GAGTTGATGAAGAAGCAGCAGCAGTAATGCTCGTCTGATCGCTTCCGGAACCTTGG 141  
::: ||||| ||| ||| ||||| ||||| :::  
Db 31 GlyIleIleGlnIuAlaIaIaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
||| ::|||  
QY 142 ATTCAGGCTACCCATGTTCTTGGCTTGACTCACGACATGGGCAATGCAATTGTGA 201  
||| ||||| ||||| :::: ||||| ::::: |||  
Db 51 IleProGlyTyrProTyrTrrPalaTrpleuGlyAspValLysTyrSerLeuSerPheThr 70  
||| ::|||  
QY 202 CGCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCTCAAGCTAAGCGCATTTCCAGAT 261  
::||| ||||| ||||| :::: |||||  
Db 71 SerArgTyrHisGluAsnSerLeuGlnLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
||| ::|||  
QY 262 GCAGCCCAAGCGGCTTGGGAATCATGGTCAACCTGGGGATGAGTGAACGGGTCGGTGGCACC 321  
||| ||||| ||| :::: ||||| ||||| |||||  
Db 91 AlaIaIaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
||| ::|||  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGGCGATTAATGTTGACACCATTTGGGGCCCGGGAAG 381  
||| ::||| ||||| ||||| |||||  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
||| ::||| ||||| ||||| |||||

[illegible]

```

RESULT 7
US-10-919-182-14
; Sequence 14, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
US-10-919-182-14

```

Alignment Scores:	
Pred. No.:	2.14e-71
Score:	774.00
Percent Similarity:	67.0%
Best Local Similarity:	47.6%
Query Match:	42.4%
DB:	6
	Gaps: 3
	Length: 369
	Matches: 150
	Conservative: 61
	Mismatches: 100
	Indels: 4

US-09-751-299-3 (1-1014) x US-10-919-182-14 (1-369)



QY 22 GCCTGCGTGAAGCCGCCCGATCTACATGATTGGAGCGGACGGTGGACAACCATT 81  
Db 11 AlaThrValGlnAlaGluProValTyrLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGAGAAGACGACGACGTAATAATGCTCGTCTGATCGCCTTCCGGAACCTGG 141  
Db 31 GlyIleIleGluGluAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCCAGGCTACCCATGGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGCAATTGTA 201  
Db 51 IleProGlyTyrProTyrTyrPalaTyrPleuGlyAspValLysTyrSerLeuSerPheThr 70  
QY 202 CGCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCTCAAGCTAAGCCGCAATTTCAGAT 261  
Db 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCCAAGCGGTTGGAAATCATGGTCAACCCCTGGGGATGATGAACGGGTCCGGTGCAC 321  
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGGCGATAATGGTGACACCATTTGGGGCCCGCGGAAG 381  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGAGTGTTCATCGCTAGCG 441  
Db 131 LeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTTCGAGACGTCGTGGGAAGCGCTGGGTGCTTATGCTGTGGAGACACCTTCAACCG 501  
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTyrGluHisIleLysGlnPro 170  
QY 502 CTAAACAATAATACGCTTTGTATGCACAAATGAAGATTCATTGTCGGCTTGGCCGAGC 561  
Db 171 LeuSerLysPheMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTyrProAla 190  
QY 562 TTTAGC--CTTATCCTAATGCGGCGGAAGCCCTGGGCTGATGTCAATGTAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTGAATCTATGCCGTGAAGGCAATGCTTCGTACTAGCGTCTGTGCGCTGTTTCA 678  
Db 210 ThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTGTGGTGT 738  
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249  
QY 739 GGTGACACTCAACGTATCATAGGCGCTGATGTGTGACTTGTCGCGCTTGGCGGAA 798  
Db 250 CysGlyTyrPalaArgIleTyrGlyProAspGlySerGluLeuAlaLysProLeuAlaGlu 269  
QY 799 AATGAAGAGGTATCTCTACGCAACCTTGATCCTGAGTACGCATCCTTGCTAAATG 858  
Db 270 AspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeuLeuAlaLysAla 289  
QY 859 GCGGACAGACCTGTGTCTATTATCCCGTCCGACATTACTCGCTTGCTAATAGATCC 918  
Db 290 GlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerValGlnPheAspPro 309  
QY 919 AGCCCTAAATTACCGGTA-----GTTGAATTGAAGTGAATCTT 957  
Db 310 ArgAsnHisThrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 8  
US-10-919-182-8  
; Sequence 8, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark

/ APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 8  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase B4  
US-10-919-182-8

Alignment Scores:  
Pred. No.: 1,81e-70 Length: 369  
Score: 765.00 Matches: 150  
Percent Similarity: 66.0% Conservative: 58  
Best Local Similarity: 47.6% Mismatches: 103  
Query Match: 41.9% Indels: 4  
DB: 6 Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-8 (1-369)

QY 22 GCCTGCGTGAAGCCGCCCGATCTACATGATTGGAGCGGACGGTGGACAACCATT 81  
Db 11 AlaThrValGlnAlaGluProValTyrLeuAspAlaAspAlaThrIleAspLysSerIle 30  
QY 82 GAGTTGATGAGAAGACGACGACGTAATAATGCTCGTCTGATCGCCTTCCGGAACCTGG 141  
Db 31 GlyIleIleGluGluAlaGlnLysGlyAlaSerLeuIleAlaPheProGluValPhe 50  
QY 142 ATTCCAGGCTACCCATGGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGCAATTGTA 201  
Db 51 IleProGlyTyrProTyrTyrPalaTyrPleuGlyAspValLysCysSerLeuSerPheThr 70  
QY 202 CGCCAATACCATGAGAACTCATTTGGAGTTGGATGGCCCTCAAGCTAAGCCGCAATTTCAGAT 261  
Db 71 SerArgTyrHisGluAsnSerLeuGluLeuGlyAspAspArgMetArgArgLeuGlnLeu 90  
QY 262 GCAGCCCAAGCGGTTGGAAATCATGGTCAACCCCTGGGGATGATGAACGGGTCCGGTGCAC 321  
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTyrSerGluArgGluAlaGlySer 110  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGGCGATAATGGTGACACCATTTGGGCGCGGAAG 381  
Db 111 ArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleValAlaAsnArgArgLys 130  
QY 382 TTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGAGTGTTCATCGCTAGCG 441  
Db 131 LeuLysProThrHisValGluArgThrIleTyrGlyGluGlyAsnGlyThrAspPheLeu 150  
QY 442 GTTTTCGAGACGTCGTGGGAAGCGCTGGGTGCTTATGCTGTGGAGACACCTTCAACCG 501  
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTyrGluHisPheGlnPro 170  
QY 502 CTAAACAATAATACGCTTTGTATGCACAAATGAAGATTCATTGTGCGGCTTGGCCGAGC 561  
Db 171 LeuSerLysIleMetMetTyrSerLeuGlyGluGlnValHisValAlaSerTyrProAla 190  
QY 562 TTTAGC--CTTATCCTAATGCGGCGGAAGCCCTGGGCTGATGTCAATGTAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGluAlaAsnAlaThrVal 209  
QY 619 TCTGAATCTATGCCGTGAAGGCAATGCTTCGTACTAGCGTCTGTGCGCTGTTTCA 678  
Db 210 IleArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229  
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTGCTGCTGTGT 738  
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeuLeuProGlnGly 249





```
OY 643 CAATGCTTCTAGCGTCG-----TGTGCG----- 669
Db 214 GLYCSPheValLeuSerAlaAsnGlnPheCysArgArgAspTyrProProPro 233
OY 670 -----CTCGTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCG 723
Db 234 GluTyrValPheAlaGlyThrGluValAspLeu-----ThrProAsp-----Ser 248
OY 724 TTGCTTCTGCTGTGTGTGGACACTCACGTCATCATAGGCGCTGATGCTGTGACTTGTC 783
Db 249 ValValCysAlaGlyGly-----SerValIleIleSerProLeuGlyAlaValLeuAla 266
OY 784 GCGCCTCTTGCCGAAATGAAGGGTATTCTCTACGCAACCTTGATCCTGGAGTACGC 843
Db 267 GlyProAsnTyrAspGly---GluAlaLeuIleSerAlaAspLeuAspLeuGlyGluIle 285
OY 844 ATCCTTGCTAAATGGCGGAGACCCCTGCTGTCAATTATCCCGTCCGACATTAATCGC 903
Db 286 AlaArgAlaLysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSer 305
OY 904 TTGCTAATAGATCGACGCCCTAAATTACCGTA 936
Db 306 LeuThrValLysAspHisProThrAsnProVal 316

RESULT 12
US-11-096-568A-6954
; Sequence 6954, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6954
; LENGTH: 350
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(350)
; OTHER INFORMATION: Ceres Seq. ID no. 15169317
US-11-096-568A-6954

Alignment Scores:
Pred. No.: 1.1e-37 length: 350
Score: 446.50 Matches: 111
Percent Similarity: 51.7% Conservative: 60
Best Local Similarity: 33.5% Mismatches: 113
Query Match: 24.5% Indels: 47
DB: 7 Gaps: 10

US-09-751-299-3 (1-1014) x US-11-096-568A-6954 (1-350)
OY 13 ATCAAGGTCGCGCTGCGCAAGCCGCCCGATCTACATGGAATTGGAGCGCAGGTGAC 72
Db 30 ValArgAlaThrValValGlnAlaSerThrIlePheTyrAspThrProAlaThrLeuAsp 49
OY 73 AAAACCATGTGAGTTGATGGAAGAAGCAGCAGCAGTAATAATGCTCTGTGATCGCCTTTCCG 132
Db 50 LysAlaGluArgLeuLeuAlaGluAlaThrSerTyrGlySerGlnLeuValValPhePro 69
OY 133 GAAACTTGATTCAGGCTACCCATGTTTCTTTGGCTTGACTCACCAGCATGGGCAATG 192
Db 70 GluAlaPheValGlyGlyTyrPro-----ArgGlySerAlaPheGlyLeu 84
OY 193 CAATTT-----GTACGCCAATACCATGAGAATCA 222
Db 85 SerIleGlyAsnArgThrValLysGlyArgGluGluPheArgLysTyrHisSerAlaAla 104
OY 223 TTGAGTTGATGGCCCTCAAGCTAAGCGCATTTCAAGTGCAGCCAAAGCGTTGGGAATC 282
```

```
Db 105 IleAspValProGlyProGluValAlaAspArgLeuAlaIleMetAlaGlyLysTyrLysVal 124
OY 283 ATGTCACCCCTGGGAGTAGTGAACGGGTCGGTGGCACCTTTACATCAGTCAGTGTTTC 342
Db 125 HisLeuValMetGlyValIleGlyLysArgAspGlyTyrThrLeuTyrCysThrValLeuPhe 144
OY 343 ATAGCGCATATGTGACACCATTTGGGGCCCGCGAAAGTTGAACCTACTTTGTGAA 402
Db 145 PheAspSerGlnGlyHisTyrLeuGlyLysHisArgLysIleMetProThrAlaLeuGlu 164
OY 403 CGTACTTTGTTGCGCGAAGGAGTGTTCATCGCTAGCGGTTTTCGAGACGTCGTTGGA 462
Db 165 ArgValIleTrpGlyPheGlyAspGlySerThrIleProValPheGluThrProValGly 184
OY 463 AGGCTGGGTGCTTATGCTGTTGGAGACACCTTCAACCGCTAACAAATACGCTTGTAT 522
Db 185 LysIleGlyAlaAlaIleCysTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyr 204
OY 523 GCACAAATGAAGATTCATTGTGCGGCTTGCGCGAGCTTTAGCCTTATCCTAATGCG 582
Db 205 AlaLysGlyValGluIleTyrCysAla-----ProThrAla 216
OY 583 GCGAAAGCCCTGGGGCTGATGTCAATGTAGCGGCTCTCGAATCTATGCGCGTTGAAGG 642
Db 217 Asp-----AlaArgAspValTrpGlnAlaSerMetThrHisIleAlaLeuGluGly 233
OY 643 CAATGCTTCTAGTCGTCG-----TGTGCG----- 669
Db 234 GLYCSPheValLeuSerAlaAsnGlnPheCysArgArgArgAspTyrProProPro 253
OY 670 -----CTCGTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCG 723
Db 254 GluTyrValPheAlaGlyThrGluValAspLeu-----ThrProAsp-----Ser 268
OY 724 TTGCTTCTGCTGTGTGTGGACACTCACGTCATCATAGGCGCTGATGCTGTGACTTGTC 783
Db 269 ValValCysAlaGlyGly-----SerValIleIleSerProLeuGlyAlaValLeuAla 286
OY 784 GCGCCTCTTGCCGAAATGAAGAGGGTATTCTTACGCAACCTTGATCCTGGAGTACGC 843
Db 287 GlyProAsnTyrAspGly---GluAlaLeuIleSerAlaAspLeuAspLeuGlyGluIle 305
OY 844 ATCCTTGCTAAATGGCGGAGACCCCTGCTGTCAATTATTCGCCGACATTAATCGC 903
Db 306 AlaArgAlaLysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSer 325
OY 904 TTGCTAATAGATCGACGCCCTAAATTACCGTA 936
Db 326 LeuThrValLysAspHisProThrAsnProVal 336

RESULT 13
US-11-096-568A-6956
; Sequence 6956, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6956
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(233)
; OTHER INFORMATION: Ceres Seq. ID no. 15169319
US-11-096-568A-6956
```



Alignment Scores:

Pred. No.: 5,21e-25 Length: 233  
Score: 323.00 Matches: 85  
Percent Similarity: 52.3% Conservative: 39  
Best Local Similarity: 35.9% Mismatches: 81  
Query Match: 17.7% Indels: 32  
DB: 7 Gaps: 8

US-09-751-299-3 (1-1014) x US-11-096-568A-6956 (1-233)

QY 265 GCCAAGCGTTGGGAATCATGTACACCTGGGGATGAGTGAACGGGTGGTGCACCCCT 324  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 2 AlaGlyLysTyrLysValHisLeuValMetGlyValIleGluArgAspGlyTyrThrLeu 21  
QY 325 TACATCAGTCAGTGTCTATAGGCGATATGTTGACACCATTTGGGGCCGGCGAAAGTTG 384  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 22 TyrCysThrValLeuPhePheAspSerGlnGlyHisTyrLeuGlyLysArgLysIle 41  
QY 385 AAACCTACTTTTGTGAACGTACTTTGTTTCGGCGAAGGGATGTTTCATCGCTAGCGGTT 444  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 42 MetProThrAlaLeuGluArgValIleTyrGlyPheGlyAspGlySerThrIleProVal 61  
QY 445 TTCGAGACGTCGTGTGGAAGCGTGGGTGCTTATGCTGTTGGGAGCACCTTCAACCGCTA 504  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 62 PheGluThrProValGlyLysIleGlyAlaIleCysTyrGluAsnArgMetProLeu 81  
QY 505 ACAAAATACGCTTGTATGCACAAATGAAGAGATTCATGTGCGGCTTGGCCGAGCTTT 564  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 82 LeuArgThrAlaMetTyrAlaLysGlyValGluIleTyrCysAla----- 96  
QY 565 AGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCGCTGATGCAATGTAGCGGCTCTCGA 624  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 97 -----ProThrAlaAsp-----AlaArgAspValTyrGlnAlaSerMetThr 110  
QY 625 ATCTATGCCGTTGAAGGGCAATGCTTCTGCTAGCGTCG-----TGTGCG----- 669  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 111 HisIleAlaLeuGluGlyGlyCysPheValLeuSerAlaAsnGlnPheCysArgArgArg 130  
QY 670 -----CTCGTTTCACAATCCATGATCGATATGCTTTGTACA 705  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 131 AspTyrProProProProGluTyrValPheAlaGlyThrGluValAspLeu-----Thr 148  
QY 706 GATGACGAAAGCATGCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 149 ProAsp-----SerValValCysAlaGlyGly-----SerValIleIleSerPro 163  
QY 766 GATGCTGTGACTTGTGTCGCGCCTCTTGCAGAAATGAAGAGGGTATTCTTACGCAAC 825  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 164 LeuGlyAlaValLeuAlaGlyProAsnTyrAspGly--GluAlaLeuIleSerAlaAsp 182  
QY 826 CTGTATCCTGAGTACGCATCCTTGTCTAAATGCGCGAGACCCCTGCTGCTATTTCC 885  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 183 LeuAspLeuGlyGluIleAlaArgAlaLysPheAspPheAspValValGlyHisTyrSer 202  
QY 886 CGTCCCGACATTAATCGCTTGTCTAATAGATCGACAGCCCTAAATTACCGGTA 936  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 203 ArgProGluValLeuSerLeuThrValLysAspHisProThrAsnProVal 219

RESULT 14

US-11-096-568A-20688  
; Sequence 20688, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; PRIOR FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20688  
; LENGTH: 193  
; TYPE: PRT

ORGANISM: Zea mays subsp. mays

; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(193)  
; OTHER INFORMATION: Ceres Seq. ID no. 12386687  
US-11-096-568A-20688

Alignment Scores:

Pred. No.: 4.92e-20 Length: 193  
Score: 274.50 Matches: 68  
Percent Similarity: 54.2% Conservative: 36  
Best Local Similarity: 35.4% Mismatches: 65  
Query Match: 15.0% Indels: 23  
DB: 7 Gaps: 7

US-09-751-299-3 (1-1014) x US-11-096-568A-20688 (1-193)

QY 388 CCTACTTTTGTGAACGTACTTGTTCGCGAAGGGATGTTTCATCGCTAGCGGTTTC 447  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 2 ProThrAlaLeuGluArgValPheTyrGlyPheGlyAspGlySerThrIleProValTyr 21  
QY 448 GAGACGTCGTGTGGAAGCGTGGGTGCTTATGCTGTTGGAGCACCTTCAACCGCTAACA 507  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 22 AspThrProIleGlyLysMetGlyAlaLeuIleCysTyrGluAsnArgMetProLeuLeu 41  
QY 508 AAATACGCTTGTATGCACAAATGAAGAGATTCATGTGCGGCTTGGCCGAGCTTTAGC 567  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 42 ArgThrAlaMetTyrAlaLysGlyIleGluIleTyrCysAla-----ProThrValAsp 59  
QY 568 CTTTATCTTAATGCGGCGAAAGCCCTGGGCGCTGATGTCATGTAGCGGCTTCGAATC 627  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 60 CysMetProThrTyr-----LeuSerSerMetThrHis 70  
QY 628 TATGCCGTTGAAGGGCAATGCTTGCNACTAGCGTGTGCTGCTGCTGCTTTCACAATCCATG 687  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 71 IleAlaLeuGluGlyGlyCysPheValLeuSerAlaCysGlnPheCysArgArgLysAsn 90  
QY 688 ATCGAT-----ATGCTTTGTACAGATGACGAAAG-----CATGCG 723  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 91 TyrProProProProGluTyrThrPheCysGlyLeuGluGluProSerProGluSer 110  
QY 724 TTGCTTCTGCTGTGTGTGACACTGACGTATCATAGGGCTGATGCTGCTGCTGCTGCTGCT 783  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 111 ValValCysSerGlyGly-----SerValIleIleSerProLeuGlyThrValLeuAla 128  
QY 784 GCGCCTCTTGGCCGAAATGAAGAGGGTATTCTTACGCAAACTTGATCCTGAGTACGC 843  
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 129 GlyProAsnTyrGluSer--GluAlaLeuLeuThrAlaAspLeuAspLeuGlyGluIle 147  
QY 844 ATCCTTGTCTAAATGCGCGGACAGACCCCTGCTGCTCATTTATCCCGCCGACATTACTCGC 903  
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 148 ValArgAlaLysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSer 167  
QY 904 TTGCTAATAGATCGACGCCCTTAATTACCGGTAAGTT 939  
||||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::  
Db 168 LeuValValLysSerAspProLys--ProAlaVal 178

RESULT 15

US-11-079-463-5457  
; Sequence 5457, Application US/11079463  
; Publication No. US20060073161A1  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRA  
; FILE REFERENCE: PATH00-03DIV2  
; CURRENT APPLICATION NUMBER: US/11/079,463  
; PRIOR FILING DATE: 2005-03-14  
; PRIOR APPLICATION NUMBER: US 60/128,705  
; PRIOR FILING DATE: 1999-04-09  
; PRIOR APPLICATION NUMBER: US 09/540,209  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:29:49 ; Search time 4775.42 Seconds  
(without alignments)  
9934.641 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaaagaagctatcaagt.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_htc:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	128	12.6	769	8	DR609823
2	113.6	11.2	840	8	DR709497
3	113	11.1	888	9	AQ329770
4	105.6	10.4	551	7	CO149108
5	105.6	10.4	553	7	CO149186
6	105.6	10.4	610	7	CO134433
7	105.6	10.4	804	8	DR709485
8	96.8	9.5	546	7	CO139118
9	93.2	9.2	840	7	CN808140
10	93.2	9.2	998	7	CN808249
11	88.4	8.7	515	7	CO137063
12	87	8.6	655	6	CF453420
13	87	8.6	699	8	DR631467
14	87	8.6	732	8	DR631867
15	87	8.6	810	8	DR629486
16	87	8.6	841	8	DR632130
17	87	8.6	856	8	DR622046
18	87	8.6	858	8	DR630522
19	86.6	8.5	771	8	DN905823
20	86.6	8.5	817	7	CK860468
21	86.4	8.5	688	6	CA265056
22	86.4	8.5	694	6	CA283378

23	86.4	8.5	840	7	CO170076	CO170076	NDL1_11_C
24	86.4	8.5	855	8	DR163242	DR163242	RTFBI_41
25	86.2	8.5	622	9	AQ329769	AQ329769	nbxb0045P
26	85.4	8.4	695	6	CA185867	CA185867	SCSFSCT307
27	84.2	8.3	810	8	DR523243	DR523243	WS02714.B
28	83.4	8.2	705	2	BG128900	BG128900	EST474546
29	83.4	8.2	738	2	BG123280	BG123280	EST468926
30	83.4	8.2	736	3	BI925888	BI925888	EST545777
31	82.6	8.1	815	7	CN136550	CN136550	OX1_43_H0
32	82.2	8.1	701	2	BE660909	BE660909	1155_Gmax
33	82.2	8.1	895	7	CO163629	CO163629	FLD1_42_G
34	81.8	8.1	607	7	CO364093	CO364093	RTK1_13_D
35	81.8	8.1	660	6	CD717845	CD717845	VVB143F06
36	81.8	8.1	668	6	CD715152	CD715152	VVB208G06
37	81.8	8.1	842	6	CB342297	CB342297	CA32EN000
38	81.6	8.0	690	3	BI418062	BI418062	LJNEST24b
39	81.6	8.0	1087	1	AW720658	AW720658	LJNEST16h
40	81	8.0	789	7	CN133019	CN133019	OX1_9_D07
41	81	8.0	792	7	CN131548	CN131548	RHOH1_48
42	81	8.0	797	7	CN124609	CN124609	RHOH1_5_G
43	81	8.0	802	7	CN126484	CN126484	RHOH1_17
44	81	8.0	833	7	CN130894	CN130894	RHOH1_44
45	80.2	7.9	902	8	DR525110	DR525110	WS02719.B

ALIGNMENTS

RESULT 1  
LOCUS DR609823 769 bp mRNA linear EST 11-JUL-2005  
DEFINITION EST999951 FVG Gibberella moniliformis CDNA clone FVGDB83, mRNA  
sequence.  
ACCESSION DR609823 GI:70684471  
VERSION DR609823.1  
KEYWORDS EST.  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Bukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
REFERENCE 1 (bases 1 to 769)  
Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Uterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.  
Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisn gene cluster  
Unpublished (2005)  
JOURNAL  
COMMENT Contact: Brown, D.W.  
USDA/ARS/NCAUR  
USDA

1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncaur.usda.gov  
TIGR sequence name: FVGDB83TH  
Seq primer: AAT TAA CCC TCA AAG GG.  
Location/Qualifiers  
1. 769

FEATURES

Source

/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="m3125"  
/db\_xref="taxon:117187"  
/clone="FVGDB83"  
/tissue\_type="mycelia"  
/clone\_lib="FVG"  
/note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Mycelia was collected after growth in liquid GYM medium for 96 hours. Cultures were vacuum filtered and the mycelial mats were frozen in liquid nitrogen, ground to a powder, and then added to TRIzol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml TRIzol. The cDNA was directionally ligated into the pBluescript II SK(+) XR



vector (CDNA Synthesis kit; Stratagene). "

Query Match	12.6%	Score 128;	DB 8;	Length 769;
Best Local Similarity	53.0%	Pred. No. 1.1e-28;		
Matches 297; Conservative	0;	Mismatches 260;	Indels 3;	Gaps 1;

QY	1	ATGAAAGAAGCTATCAAGGTCGCTGCGTCAAGCCGCCCGATCTACATGATTTGGAG	60
Db	43	ATGTCCAAGTCCCTCAAGGTGCGCCCATCCAAGCCGAGCCCGCTTGGAACGATCTCCAG	102
QY	61	GCGACGGTGGACAAAACCATGTAGTTGATGGAAGAAGCAGCACGTAATAAGTCTGCTG	120
Db	103	GGTGGCGTCAACAAGTCCATCGGTCTCATCCAAGAGCAGCAAGAGGCGTCCAAAGTC	162
QY	121	ATCGCCTTTCGGAAACTTGGATTCCAGGCTACCCATGTTCTTTGGCTGACTCACC	180
Db	163	ATTGGCTTCCCTGAAGTCTTCATCCCTGGATATCCATGGAGCATCTGGGCCA	222
QY	181	GCATGGGCAATGCAATTGTGTACGCCAATACCATGAGAATCATTTGAGTTGATGGCCCT	240
Db	223	ACCGAGAACGCCACCATGGATCAATGAGTACTTCAAGAATCAATGAGAAAGATCACCT	282
QY	241	CAAGCTAAGCGCATTTCAGATGACGCCAAGCGTTGGGAATCATGTCAACCTGGGGATG	300
Db	283	GAGATGGAACCATGATCCGAGCTGCTGTTCGAGAGGACAGTGTCTTTGTAGTCTTGATAC	342
QY	301	AGTGAACGGGTCCGTGGTGCACCCCTTTACATCAGTGTTCATATAGGCGATAATGTGAC	360
Db	343	AGCGAGAGATACAGGGGAACCTTTACATCGCACAGTCTTTCATCGACGACACCGGCACT	402
QY	361	ACCATTGGGGCCCCGGCGAAAGTTGAAACTTACTTTTGTGAACGTACTTTGTTCCGGCA	420
Db	403	ATTGTTCTCCACCGTCCGAAGATCAAGCTTACCCACGTTGAGCGTGTCTATCTACGGTGAC	462
QY	421	GCGGATGGTTCATCGCTAGC--GGTTTCGAGACGTCTGTTGGAAGCGTGGGTGGCTTA	477
Db	463	GGACAGGGCGAGTCTCTGACCAACGTGCGCGACACAAGATTGGAAGGTTGCTGTTCTT	522
QY	478	TGCTGTGGGAGCACTTCAACCGCTAACAATAACGCTTTGTATGACAAATGAAGAG	537
Db	523	AACTGCTGGAGACACACCCAGACACTTCTCCGCTACTACGAATACTCCCAAGACGTGAT	582
QY	538	ATTCATTGTGCGGCTTGCC	557
Db	583	ATCCACGTCTCCAGCTGGCC	602

RESULT 2  
DR709497  
LOCUS  
DEFINITION  
Asn\_11063 *Aspergillus niger* pBluescript (EcoRI-XhoI) *Aspergillus niger* cDNA clone Asn\_11063, mRNA sequence.  
840 bp mRNA linear EST 14-JUL-2005  
ACCESSION  
DR709497  
VERSION  
DR709497.1 GI:70825788  
KEYWORDS  
EST.  
SOURCE  
*Aspergillus niger*  
ORGANISM  
*Aspergillus niger*  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.  
1 (bases 1 to 840)  
REFERENCE  
Tsang, A., Storms, R. and Bulter, G.  
AUTHORS  
Expressed sequence tags from *Aspergillus niger* cDNA library  
TITLE  
Unpublished (2005)  
JOURNAL  
Contact: Tsang A  
COMMENT

## FEATURES

Location/Qualifiers

**Source**

```

1. .840
/organism="Aspergillus niger"
/mol_type="mRNA"
/strain="N402"
/db_xref="taxon:5061"
/clone="Asn_11063"
/dev_stage="mycelial growth"
/lab_host="E. coli"
/clone_lib="Aspergillus niger pluescript (EcoRI-XhoI)"
/note="Vector: pluescript KS+; Site 1: XhoI; Site 2:
EcoRI; Complementary DNA was synthesized with ZAP Kit
(Stratagene) using poly(A)+RNA isolated from Aspergillus
niger cultured under different carbon sources (glucose,
maltose, xylose, lactose, sorbitol, xylan, and bran).
Synthesis was primed with oligo(dT)/XhoI primer. EcoRI
adaptors were ligated to the blunt-ended, double-stranded
cDNA. The EcoRI-XhoI-digested cDNA was ligated with
EcoRI-XhoI-digested pluescript KS+ (Invitrogen Corp)."
```

## ORIGIN

Query Match	11.2%;	Score 113.6;	DB 8;	Length 840;
Best Local Similarity	49.6%;	Pred. No. 4.1e-24;		
Matches 411;	Conservative 0;	Mismatches 399;	Indels 18;	Gaps 4;

QY	19	GTCCGCTGCGTGCAAGCCGCCCCGATCTACATGGAATTTGGAGGCCGCGGTGGACAAAACC	78
Db	22	GTCGCAGTCAACCAGGAGAGCCAGTCTGGCTGGACTTTGGAAACAACCGTGAAAAAGACA	81
QY	79	ATTGAGTTGATGGAAGACGACGACGTAATATGCTGCTGATGCGCTTCCGGAACCT	138
Db	82	TGTGACCTTATTGCTGAAGCAGCTGCGAACGCGCTCAGTTGCTGACTTTTCCCGAGTGT	141
QY	139	TGGATTCAGGCTAACCAATGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGCMAATTT	198
Db	142	TGGATCCCGGATATCTGCTCGATTGG-----GCACGCGCTGTTGACATGCGCCTA	195
QY	199	GTACGCCAATACCATGAGAACTCATTTGAGTTGATGCGCTCAAGCTTAAGCGCATTTCA	258
Db	196	TCATCTATTACATACAAAACCTCCCTGAAGATTGATTCGCCGCAATGGCCAGCATCCAG	255
QY	259	GATGCAGCCGAAGCGGTTGGGAATCATGATCACCCTGGGGAATGATGAACGGTGGTGAC	318
Db	256	CAATGCCGACGGAGAAATAAATCGTGTTGTGTGGGCTTTCGAGAACTGCATAAC	315
QY	319	ACCCTTTACATCAGTCAGTGTTCATTAGGCGATATGTTGACACCATTTGGGGCCCCGCGA	378
Db	316	TCCCTGTATATCTCGCAGGCTATATTTCGAAGTGATGGGAGATCTCACCACCCGCAAG	375
QY	379	AAGTTGAAACCTACTTTTGTGAAACGTACTTGTTCGGCGGAAGGGAATGTTTCATCGCTA	438
Db	376	AAATCAAGCCCACTCATATGAGCGGACCAATTTTCGGTGACTCGTTTGGAGACTGTCTG	435
QY	439	GCG---GTTTTCGAGACGTCGTGTTGGAAGGCTGGGTGCTTATGCTGTTGGAGCACTT	495
Db	436	CAGAGTGTAGTCGATACATCAGCCGGTCGCGTTGTGCTCTTTCCTGCTGGGAGCATATC	495
QY	496	CAACCGCTAACAAAATACGCTTTGTATGCACAAAATGAAGAGATTCAATTGTCGGCTTG	555
Db	496	CAGCCTCTGCTCAAGTATCACACTTATGCGCAGCGGGAACAATTCATGTTGCCGCATGG	555
QY	556	CCGAGCTTTAGCCTTTATCTTAA-----TCCGCGGAAAAGCCCTGGGCGCTGATGTCAAT	609
Db	556	CCGCTCTGTTCCTCATAGCGAGGACGGTTCCTGTTTTCGATGTCCACCGAGGGCACC	615
QY	610	GTAAGCGGCTCTCGAATCTATGCGCTGAAGGGCAATGCTTGTAAGCGTCTGTGCG	669
Db	616	AGTTTCATTGCGAGACCTATGCCATCGAAAGTCAGTCTTTTGTCTCCATACACAACA	675
QY	670	CTCGTTTACAATCCATGATCGATATGCTTTGTACAGATGACGAAGAAGCATGCTTGCTT	729
Db	676	GTCATAGGCCAGTCTGGCATTTGACCGGATGGCCACATCGACTGGCGCCC---TGATGAGC	732
QY	730	CTGGCTGTGTGTGGACATCACGTAATCATAGGGCGCTGATGTTGTGTGACTTTGTCGGCGCT	789

Db 733 ACCCTGGCGAGGTGCTCCGCCATCTTCGGGCCCCGACGGTGCCCAATATTCGCAACC 792  
Oy 790 CTGCGGAAATGAGAGGGTATTTCTACGCAACCTTGATCCTGA 837  
Db 793 ATTCTAATGACAGAGGGATTATCTACCCCGACTTGACTTTGAA 840

RESULT 3

AQ329770/c

LOCUS 888 bp DNA linear GSS 08-JAN-1999

DEFINITION

nbxb0045P09r CUGI Rice BAC Library Oryza sativa (japonica cultivar-group) genomic clone nbxb0045P09r, genomic survey sequence.

AQ329770

ACCESSION AQ329770.1 GI:4121620

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaeae; Oryza.

1 (bases 1 to 888)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

CONTACT: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAACACGTATGACCATG

Class: BAC ends

High quality sequence start: 13

High quality sequence stop: 225.

Location/Qualifiers

1. 888

/organism="Oryza sativa (japonica cultivar-group)"

/mol\_type="genomic DNA"

/cultivar="japonica"

/cultivar="Nipponbare"

/db\_xref="taxon:39947"

/clone="nbxb0045P09r"

/tissue\_type="leaf"

/lab\_host="E. coli DH10B"

/clone\_lib="CUGI Rice BAC Library"

/note="Vector: pBelobAC1; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

ORIGIN

Query Match 11.1%; Score 113; DB 9; Length 888;

Best Local Similarity 56.2%; Pred. No. 6.5e-24;

Matches 209; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Oy 183 ATGGCAATGCAATTGTGTACGCCAATACATGAGAACTCATTTGAGTTGATGCCCCCTCA 242  
Db 372 ATGGGGAAGCAGTTAGACGATCGACAGTCAAGAAAACAGCANGNCCGAGATGCCCCGA 313  
Oy 243 AGCTAAGCGCATTCAGATGACGCCAAGCGTTGGAAATCATGTTCACCCCTGGAGTAG 302  
Db 312 GTCCGCCGNATCGCCGAGGCCCGCGCTCGGCAAGGCATCTGATCAGTCTGGCTATAG 253  
Oy 303 TGAACGGGTCCGTGACCCCTTTACATCAGTCACTGTTTCATAGGCGATAATGTTGACAC 362  
Db 252 CGAGCGCAGCGCCCGGACGCTCTACATCGCTCAACTACTGATGATCGTGGCAGCAT 193  
Oy 363 CATTGGGGCCCGGGAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAAGG 422  
Db 192 CGTACAGCGCGCAGCGCAAGCTCAAGGCAACCCATGTGAGCGGACCATCTTGGTGAAGG 133  
Oy 423 GGATGTTTCATCGCTAGCGGTTTTCGAGACGCTGTGGAAGGCTGGGCTTATGCTG 482  
Db 132 CGACGCGACGACATCGCCGCTGTGAGCGACCTTGCGCGCTTGATCGCTTGCTG 73  
Oy 483 TTGGAGCACCTTCAACCGCTAACAATAACGCTTGTATGACACAAATGAGAGATTCA 542  
Db 72 CTGGGAACATATGAACCCGCTCACCAATATGCCATGTACGCCCAGACGACAGATCCA 13  
Oy 543 TTGTGCGGCTTG 554  
Db 12 TCTCGCGGCATG 1

RESULT 4

COL49108

LOCUS

DEFINITION

EST824161 Aspergillus flavus Normalized cDNA Expression Library

Aspergillus flavus cDNA clone NAGDP49 5' end, mRNA sequence.

COL49108

ACCESSION COL49108.1 GI:48903109

VERSION

KEYWORDS

SOURCE

ORGANISM

Aspergillus flavus

Aspergillus flavus

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

1 (bases 1 to 551)

Yu, J., Whitelaw, C.A., Niernan, W.C., Bhatnagar, D. and Cleveland, T.E. Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops

FEMS Microbiol. Lett. (2004) In press

CONTACT: Yu J

Food and Feed Safety Research Unit

USDA/ARS, Southern Regional Research Center

1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA

Tel: 504 286 4405

Fax: 504 286 4419

Email: jiyu@src.ars.usda.gov

Contact Dr. Yu at USDA/ARS SRRC (jiyu@src.ars.usda.gov) for clone information

PCR Primers

FORWARD: M13F

BACKWARD: M13R

Seq primer: M13 Forward

POLYA=No.

Location/Qualifiers

1. 551

/organism="Aspergillus flavus"

/mol\_type="mRNA"

/strain="NRRL 3357"

/db\_xref="taxon:5059"

/clone="NAGDP49"

/sex="asexual mycelia"

/cell\_type="mycelia"

/dev\_stage="developmental stages from 18 to 96 hours"

/lab\_host="E. coli DH10B T1 resistant cells"

/clone\_lib="Aspergillus flavus Normalized cDNA Expression Library"

/note="Vector: pBluescript (SK+) (Stratagene), antibiotic selection marker: Carbenicillin; Site\_1: NotI, at the 5 prime end; Site\_2: EcoRI, at the 3 prime end; This normalized cDNA\_expression library was constructed using a mixture of mycelial cells grown under eight different medium conditions and harvested at 5 time points (18, 24, 48, 72, 96 hours). The poly-A sequence was trimmed off before ligating to vector."

ORIGIN

Query Match	10.4%;	Score 105.6;	DB 7;	Length 551;
Best Local Similarity	51.5%;	Pred. No. 1.3e-21;		
Matches	268;	Conservative	0;	Mismatches 249; Indels 3; Gaps 1;
QY	14	TCAAGTGCCTGCGTCAAGCCGCCGATCTACATGATTGGAGCGACGGTGACA	73	
Db	32	TCAAGTGTCTGCGTTCAAGCTGAGCCTGTCTGGAATGACCTACAAGCGGGTTCCA	91	
QY	74	AAACCATTTGATTGATGAGAAGACGACGCTAATAATGCTCTGATCGCCTTCCGG	133	
Db	92	AGGTATCTCACTGATCAAGAGACGCTGCCAAAAACGGGCCAAGTGTGGTTCCCG	151	
QY	134	AAACTTGATTCAGGCTACCCATCCATGTTCTTGGCTTGACTCACCATGGCAATGC	193	
Db	152	AGGTATTCATCCCTGGTTATCCATGAGTATCTGACTAATCGCTATTGAGATGTTT	211	
QY	194	AATTTGACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCA	253	
Db	212	CCTTCATGAACGAGTACTTCGAGAACTCCATGGAAGGAATCGGAGAGATGGAACGCA	271	
QY	254	TTTCAGATGCAGCCCAAGCGGTGGGAATCATGTCAACCTGGGGATGATGACGGGTCG	313	
Db	272	TCAAAATTGCCGTGCGGCAAGCGGGTGATTCATCGTTCTAGGGTATAGTGAACGTACC	331	
QY	314	GTGGCACCCCTTACATCATGAGTGTTCATAGCGCATTAATGTTGACACCATTTGGGGCC	373	
Db	332	GAGGATCATGTGTACTGTCTCAGTCATTATCGATCCTACTGTACTATCTGCCACCACC	391	
QY	374	GGCGAAAGTTGAAACCTATTGTTGTAACGTAATTGTTCCGCCGAAGGGGATGTTTCA	433	
Db	392	GACGGAAGATCAAGCCGATCATGTGAGAGAACTATTGGGGAGACGGGCGAGCAACT	451	
QY	434	CGCT--AGCGTTTTCAGACGCTCTGTTGGAAGCTGGGTGGCTTATGCTTTGGGAGC	490	
Db	452	CTCTTAAGACCGGTAGCCCCCAGTGCTTTTGGAAACATTGGGGGGCTCAACTGTTGGAGC	511	
QY	491	ACCTTCAACCGCTAACAAATAGCGTTTGTATGCACAAAA	530	
Db	512	ATACCCAACCACTTCTACGATATTACGAGTATGCTCAAAA	551	
RESULT 5				
COL49186		553 bp	mRNA	linear
LOCUS				EST 17-JUN-2004
DEFINITION		EST824239 Aspergillus flavus Normalized cDNA Expression Library		
ACCESSION		COL49186		
VERSION		COL49186.1		GI:48903187
KEYWORDS		EST.		
SOURCE		Aspergillus flavus		
ORGANISM		Aspergillus flavus		
REFERENCE		1 (bases 1 to 553)		
AUTHORS		Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.		
TITLE		Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops		
JOURNAL		FEMS Microbiol. Lett. (2004) In press		
COMMENT		Contact: Yu J		
		Food and Feed Safety Research Unit		
		USDA/ARS, Southern Regional Research Center		
		1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA		
		Tel: 504 286 4405		

Fax: 504 286 4419  
Email: jiyu@srcc.ars.usda.gov  
Contact Dr. Yu at USDA/ARS SRRC (jiyu@srcc.ars.usda.gov) for clone information  
PCR Primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.

FEATURES

source

Location/Qualifiers  
1..553  
/organism="Aspergillus flavus"  
/mol\_type="mRNA"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAGDQ49"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="E. coli DH10B T1 resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression Library"  
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic selection marker: Carbenicillin; Site\_1: NotI, at the 5 prime end; Site\_2: EcoRI, at the 3 prime end; This normalized cDNA\_expression library was constructed using a mixture of mycelial cells grown under eight different medium conditions and harvested at 5 time points (18, 24, 48, 72, 96 hours). The poly-A sequence was trimmed off before ligating to vector."

ORIGIN

Query Match	10.4%;	Score 105.6;	DB 7;	Length 553;
Best Local Similarity	51.5%;	Pred. No. 1.3e-21;		
Matches	268;	Conservative	0;	Mismatches 249; Indels 3; Gaps 1;
QY	14	TCAAGTGCCTGCGTCAAGCCGCCGATCTACATGATTGGAGCGACGGTGACA	73	
Db	34	TCAAGTGTCTGCGTTCAAGCTGAGCCTGTCTGGAATGACCTACAAGCGGGTTCCA	93	
QY	74	AAACCATTTGATTGATGAGAAGACGACGCTAATAATGCTCTGCTGATCGCCTTCCGG	133	
Db	94	AGTCATCTCACTGATCAAGGACGCTGCCAAAAACGGGCCAAGGTGTCGGTTTCCCG	153	
QY	134	AACTTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCACCATGGGCAATGC	193	
Db	154	AGGTATTCATCCCTGGTTATCCATGAGTATCTGACTAATCTCGGCTATTGAGAAATGTTT	213	
QY	194	AATTTGACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCA	253	
Db	214	CCTTCATGAACGAGTACTTCGAGAACTCCATGGAAGGAATCGGAGAGATGGAACGCA	273	
QY	254	TTTCAGATGCAGCCAAAGCGGTGGGAATCATGTCAACCTGGGGATGATGTAACGGGTCG	313	
Db	274	TCAAAATTGCCGTGCGGCAAGCGGGTGATTCATCGTTCTAGGGTATAGTGAACGGTACC	333	
QY	314	GTGGCACCCCTTACATCATGAGTGTTCATAGCGCATTAATGTTGACACCATTTGGGGCC	373	
Db	334	GAGGATCATGTGTACTGTCTCAGTCATTATCGATCCTACTGTACTATCTGCCACCACC	393	
QY	374	GGCGAAAGTTGAAACCTACTTTGTTGAACGTAATTGTTCCGCCGAAGGGGATGTTTCA	433	
Db	394	GACGGAAGATCAAGCCGATCATGTGAGAGAACTATTGGGAGACGGGCGAGCAAGACT	453	
QY	434	CGCT--AGCGTTTTCAGACGCTCTGTTGGAAGCTGGGTGGCTTATGCTTTGGGAGC	490	
Db	454	CTCTTAAGACCGGTAGCCCCCAGTGCTTTTGGAAACATTGGGGGCTCAACTGTTGGAGC	513	
QY	491	ACCTTCAACCGCTAACAAATAGCGTTTGTATGCACAAAA	530	
Db	514	ATACCCAACCACTTCTACGATATTACGAGTATGCTCAAAA	553	

```
RESULT 6
COL34433      610 bp      mRNA      linear      EST 17-JUN-2004
LOCUS          Aspergillus flavus Normalized cDNA Expression Library
DEFINITION    Aspergillus flavus cDNA clone NAFAS28 5' end, mRNA sequence.
ACCESSION     COI34433
VERSION       COI34433.1  GI:48883411
KEYWORDS      EST.
SOURCE        Aspergillus flavus
ORGANISM      Aspergillus flavus
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
              Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE     1 (bases 1 to 610)
AUTHORS       Yu,J., Whitelaw,C.A., Nierman,W.C., Bhatnagar,D. and Cleveland,T.E.
TITLE         Aspergillus flavus expressed sequence tags for identification of
              genes with putative roles in aflatoxin contamination of crops
JOURNAL       FEMS Microbiol. Lett. (2004) In press
COMMENT       Contact: Yu J
              Food and Feed Safety Research Unit
              USDA/ARS, Southern Regional Research Center
              1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA
              Tel: 504 286 4405
              Fax: 504 286 4419
              Email: jiuyu@rrc.ars.usda.gov
              Contact Dr. Yu at USDA/ARS SRRC (jiuyu@rrc.ars.usda.gov) for clone
              information
PCR PRIMERS   FORWARD: M13F
              BACKWARD: M13R
              Seq primer: M13 Forward
              POLYA=No.

FEATURES
  source       1..610
                /organism="Aspergillus flavus"
                /mol_type="mRNA"
                /strain="NRRL 3357"
                /db_xref="taxon:5059"
                /clone="NAFAS28"
                /sex="asexual mycelia"
                /cell_type="mycelia"
                /dev_stage="developmental stages from 18 to 96 hours"
                /lab_host="E. coli DH10B T1 resistant cells"
                /clone_lib="Aspergillus flavus Normalized cDNA Expression
                library"
                /note="Vector: pBluescript (SK+) (Stratagene), antibiotic
                selection marker: Carbenicillin; Site_1: NotI, at the 5
                prime end; Site_2: EcoRI, at the 3 prime end; This
                normalized cDNA expression library was constructed using a
                mixture of mycelial cells grown under eight different
                medium conditions and harvested at 5 time points (18, 24,
                48, 72, 96 hours). The poly-A sequence was trimmed off
                before ligating to vector."

ORIGIN
Query Match      10.4%; Score 105.6; DB 7; Length 610;
Best Local Similarity 51.5%; Pred. No. 1.3e-21;
Matches 268; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 14 TCAGAGTCGCGCTGCGTGAAGCCGCCCGATCTACATGATTGGAGGCGACGGTGACA 73
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  91 TCAAGTTGCTGCGGTCAAGCTGAGCCCTGTCTGGAATGACCTACAAGCGGGTTTCCA 150
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 74 AAACCATGTGAGTGAAGAAGACGACAGCTAATAATGCTGCTGATGCGCTTTCCGG 133
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  151 AGGTCACTCTACTGATCAAGACGCTGCCAAAACGGCGCCAAGGTGTTGCCCG 210
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 134 AAAGTTGATTCAGGCTACCACTAGTTTCTTTGGCTTGACTCACCGATGGCAATGC 193
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  211 AGGTATTCATCCCTGTTATCCATGAGATATCTGACTAATCGGCTATTGAGATGTTT 270
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 194 AATTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCA 253
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  271 CCTTCATGAACGAGTACTTGCAGAACTCCATGGAAGGAATCGAGAGATGAACGCA 330
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

```
QY 254 TTTCAGATGCAGCCCAAGCGGTTGGGAATCATGTACACCTGGGGATGATGAACGGGTGC 313
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  331 TCAAAATTGCCGTGCGCGGAAGCGGTGTATTATTCATCGTTCTAGGGTATAGTAACGGTACC 390
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 314 GTGGCAACCTTTACATCAGTCAGTGTTCATAGGCGAATAATGTGACACCATTTGGGCC 373
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  391 GAGGATCATTTGTACATTTGCTCAGTCATTATTCATCTACTGTACTATCTGTCACGACC 450
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 374 GCGGAAAGTTGAACCTACTTTGTTGAACGTACTTGTTCGGCGAAGGGATGTTTCAT 433
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  451 GACGGAAGATCAAGCCGACTCATGTGAGAGAAGCTATTGGGAGACGTGCAAGCAAGACT 510
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 434 CGCT--AGCGGTTTTCGAGACGTCTGTTGGAAGGCTGGGTGCTTATGCTGTTGGAGC 490
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  511 CTCTTAAGACCGGTAGCCCCCAGTGCTTTTGGAACATTGGGGGGCTCAACTGTTGGGAGC 570
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 491 ACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAA 530
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  571 ATACCCAACCACTTTCAGATATTACGAGTATGCTCAAA 610
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 7
DR709485      804 bp      mRNA      linear      EST 14-JUL-2005
LOCUS          Asn_11049 Aspergillus niger pBluescript (EcoRI-XhoI) Aspergillus
DEFINITION    niger cDNA clone Asn_11049, mRNA sequence.
ACCESSION     DR709485
VERSION       DR709485.1  GI:70825776
KEYWORDS      EST.
SOURCE        Aspergillus niger
ORGANISM      Aspergillus niger
              Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
              Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
REFERENCE     1 (bases 1 to 804)
AUTHORS       Tsang,A., Storms,R. and Bulter,G.
TITLE         Expressed sequence tags from Aspergillus niger cDNA library
JOURNAL       Unpublished (2005)
COMMENT       Contact: Tsang A
              Centre for Structural and Functional Genomics
              Concordia University
              7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada
              Tel: 514 848 2424 3405
              Fax: 514 848 4504
              Email: tsang@vax2.concordia.ca
              POLYA=No.

FEATURES
  source       1..804
                /organism="Aspergillus niger"
                /mol_type="mRNA"
                /strain="N402"
                /db_xref="taxon:5061"
                /clone="Asn_11049"
                /dev_stage="mycelial growth"
                /lab_host="E. coli"
                /clone_lib="Aspergillus niger pBluescript (EcoRI-XhoI)"
                /note="Vector: pBluescript KS+; Site_1: XhoI; Site_2:
                EcoRI; Complementary DNA was synthesized with ZAP kit
                (Stratagene) using poly(A)+RNA isolated from Aspergillus
                niger cultured under different carbon sources (glucose,
                maltose, xylose, lactose, sorbitol, xylan, and bran).
                Synthesis was primed with oligo(dT)/XhoI primer. EcoRI
                adaptors were ligated to the blunt-ended, double-stranded
                cDNA. The EcoRI-XhoI-digested cDNA was ligated with
                EcoRI-XhoI-digested pBluescript KS+ (Invitrogen Corp)."
```

```
ORIGIN
Query Match      10.4%; Score 105.6; DB 8; Length 804;
Best Local Similarity 52.0%; Pred. No. 1.4e-21;
Matches 290; Conservative 0; Mismatches 259; Indels 9; Gaps 2;

QY 19 GTGCGCTGCGTGAAGCCGCCCGATCTACATGATTGGAGGCGACGGTGACAAACC 78
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```



Db 23 GTCGAGTCAACCCAGGAGAGCCAGTCTGGCTGGACTTGGAGCAACCGTGAAGAAAGACA 82

QY 79 ATTGAGTTGATGGAAGAAGCAGCAGCTAATAATGCTCGTCTGATCGCTTTCCGGAAC 138

Db 83 TGTGACCTTATGCTGAAGCAGCTGCGAAGCGCGCTCAGTTGGTGACTTTCCCGAGTGT 142

QY 139 TGGATTCCAGGCTAACCCATGTTTCTTTGGCTTGAATCACCAGCAGTGGCAATGCAATT 198

Db 143 TGGATCCCGGATATCC-----TGCTGATTTGGGCACGGCGCTGTTGACATGCGCCTA 196

QY 199 GTACGCCAATACATGAGACTCATTTGAGTTGGATGGCCCTCAAGCTTAAGCGCATTTCA 258

Db 197 TCATCTATTACATACAAACTCCCTGAAGATTGATCCCGCAATGGCCAGCATCCAG 256

QY 259 GATGACGCCAAGCGGTGGGAATCATGTCAACCTGGGGATGAGTGAACGGGTCGGTGGC 318

Db 257 CAATGCCGACGGAGATAAATCGTGTGTTGTTGGGCTTTTCCGAGAACTGCATTAAC 316

QY 319 ACCCTTTACATCAGTCACTGTTTATAGGCGATTAATGTCACACCAATTGGGGCCCGCGA 378

Db 317 TCCCTGATATATCGCAGGCTATTATTGCAAGTTATGGGAAGATCCTCACCACCCGCAAG 376

QY 379 AAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGATGTTCAATCGCTA 438

Db 377 AAAATCAAGCCCACTCATATGAGCGGACCATTTTCGTGACTCGTTGGAGACTGTCTG 436

QY 439 GCG---GTTTCGAGAGCTCTGTGGAAGGCTGGGCTTATGCTGTTGGAGACCTT 495

Db 437 CAGAGTGAATGATACATACATCAGCGGCTCGCGTTGTGCTCTTCTGCTGGAGCATATC 496

QY 496 CAACCGCTAACAAATACGCTTGTATGACAAATGAAGAGATTCAATGTGCGGCTGG 555

Db 497 CAGCCTCTGCTCAAGTATCACACTTATGCGCAGCGGAACAAATTCAATGTTGCCGATGG 556

QY 556 CCGAGCTTACGCTTTAT 573

Db 557 CCGCCTCTGTTCCCTCAT 574

RESULT 8

LOCUS COL39118 546 bp mRNA linear EST 17-JUN-2004

DEFINITION EST833789 Aspergillus flavus Normalized cDNA Expression Library

ACCESSION COL39118

VERSION COL39118.1 GI:48888096

KEYWORDS EST.

SOURCE Aspergillus flavus

ORGANISM Aspergillus flavus

REFERENCE 1 (bases 1 to 546)

AUTHORS Yu,J., Whitelaw,C.A., Niernan,W.C., Bhatnagar,D. and Cleveland,T.E.

TITLE Aspergillus flavus expressed sequence tags for identification of genes with putative roles in aflatoxin contamination of crops

JOURNAL FEMS Microbiol. Lett. (2004) in press

COMMENT Contact: Yu J

Food and Feed Safety Research Unit

USDA/ARS, Southern Regional Research Center

1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA

Tel: 504 286 4405

Fax: 504 286 4419

Email: jiu@usrrc.ars.usda.gov

Contact Dr. Yu at USDA/ARS SRRC (jiu@usrrc.ars.usda.gov) for clone information

PCR Primers

FORWARD: M13F

BACKWARD: M13R

Seq primer: M13 Forward

POLYA=No.

FEATURES

Source Location/Qualifiers

1..546

/organism="Aspergillus flavus"

/mol\_type="mRNA"

/strain="NRRL 3357"

/db\_xref="taxon:5059"

/clone="NAFDB72"

/sex="asexual mycelia"

/cell\_type="mycelia"

/dev\_stage="developmental stages from 18 to 96 hours"

/lab\_host="E. coli DH10B T1 resistant cells"

/clone\_lib="Aspergillus flavus Normalized cDNA Expression Library"

/note="vector: pBluescript (SK+) (Stratagene), antibiotic selection marker: Carbenicillin; Site\_1: NotI, at the 5 prime end; Site\_2: EcoRI, at the 3 prime end; This normalized cDNA expression library was constructed using a mixture of mycelial cells grown under eight different medium conditions and harvested at 5 time points (18, 24, 48, 72, 96 hours). The poly-A sequence was trimmed off before ligating to vector."

ORIGIN

Query Match 9.5%; Score 96.8; DB 7; Length 546;

Best local Similarity 51.2%; Pred. No. 7.8e-19;

Matches 252; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

QY 15 CAAGTGCCTCGTGCAGCGCCCGATCTACATGATTGGAGCGACGTTGACAA 74

Db 48 CAAGTTGCTGCGGTTCAAGCTGAGCCTGTCTGGAATGACCTACAAGCGGGTTCCAA 107

QY 75 AACCATGAGTTGATGGAAGACGACGTAATAATGCTGCTGATCGCCTTCCGA 134

Db 108 GGTCACTCACTGATCAAGGACGCTGCCAAAAACGGCCCAAGGTGTCGTTCCCGA 167

QY 135 AACTGGAATCCAGGCTACCCATGTTCTTTGGCTTGACTCAACGATGGGCAATGCA 194

Db 168 GGTATTCACTCCCTGTTATCCATGAGATCTGACTAATCTGGCTATTGAGATGTTT 227

QY 195 ATTGTAGCCCATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCAT 254

Db 228 CTTCAATGAACGAGTACTTGCAGAACTCCATGAAAGGGAATCGAGAGATGGAACGCAT 287

QY 255 TTCAGATGACGCCAAGCGGTTGGAAATCATGTCACCCCTGGGATGATGAACGGGTCCG 314

Db 288 CAAATATGCCGTGCGGGAAGCGGTTGATTCATGTTCTAGGTTATGTAACGGTACC 347

QY 315 TGGCACCTTTACATCAGTCAGTGTGTTCAATAGCGATATGTTGGGCCCG 374

Db 348 AGGATCATTTGATCTGCTCAGTCATTTATGATCTTACTGTACTATTCGTCCACCACCG 407

QY 375 GCGAAAGTTGAACCTACTTTTGTGAACGTAATTGTTGCGGCAAGGGATGTTCAATC 434

Db 408 ACGAAGATCAAGCCGACTCATGTGAGAGAGCTATTGGGAGACGTGCAAGCAGACTC 467

QY 435 GCT--AGCGTTTTCAGAGAGCTGTGTTGAAGCGTGGGTGCTTATGCTGTTGGAGCA 491

Db 468 TCTTAAGACGTTAGCCCCAGTGTCTTTGGAACATTGGGGGCTCAACTGTTGGAGCA 527

QY 492 CCTCAACCGCT 503

Db 528 TACCAACCACT 539

RESULT 9

LOCUS CN808140 840 bp mRNA linear EST 27-MAY-2004

DEFINITION Blood EST0223 Metarhizium anisopliae ARSEF 2575 from insect blood

ACCESSION CN808140

VERSION CN808140.1 GI:47729613

KEYWORDS EST.

SOURCE Metarhizium anisopliae

ORGANISM Metarhizium anisopliae

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic



QY	552	TTGGCC	557	
DB	648	ATGGCC	653	
RESULT 11				
LOCUS	COL137063	515 bp	mRNA	linear EST 17-JUN-2004
DEFINITION	EST831734 Aspergillus flavus Normalized cDNA Expression Library			
ACCESSION	Aspergillus flavus cDNA clone NAFC28 5' end, mRNA sequence.			
VERSION	COL137063			
KEYWORDS	COL137063.1	GI:48886041		
SOURCE	EST.			
ORGANISM	Aspergillus flavus			
	Aspergillus flavus			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;			
	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.			
REFERENCE	1 (bases 1 to 515)			
AUTHORS	Yu,J., Whitelaw,C.A., Nierman,W.C., Bhattacharjee,D. and Cleveland,T.E.			
TITLE	Aspergillus flavus expressed sequence tags for identification of			
	genes with putative roles in aflatoxin contamination of crops			
JOURNAL	FEMS Microbiol. Lett. (2004) In press			
COMMENT	Contact: Yu J			
	Food and Feed Safety Research Unit			
	USDA/ARS, Southern Regional Research Center			
	1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA			
	Tel: 504 286 4405			
	Fax: 504 286 4419			
	Email: jiyu@srcc.ars.usda.gov			
	Contact Dr. Yu at USDA/ARS SRRC (jiyu@srcc.ars.usda.gov) for clone			
	information			
	PCR Primers			
	FORWARD: M13F			
	BACKWARD: M13R			
	Seq primer: M13 Forward			
	POLYA=NO.			
FEATURES	Location/Qualifiers			
Source	1. 515			
	/organism="Aspergillus flavus"			
	/mol_type="mRNA"			
	/strain="NRRL 3357"			
	/db_xref="taxon:5059"			
	/clone="NAFC28"			
	/sex="asexual mycelia"			
	/cell_type="mycelia"			
	/dev_stage="developmental stages from 18 to 96 hours"			
	/lab_host="E. coli DH10B T1 resistant cells"			
	/clone_lib="Aspergillus flavus Normalized cDNA Expression			
	Library"			
	/note="Vector: pBluescript (SK+) (Stratagene), antibiotic			
	selection marker: Carbenicillin; Site_1: NotI, at the 5			
	prime end; Site_2: EcoRI, at the 3 prime end; This			
	normalized cDNA expression library was constructed using a			
	mixture of mycelial cells grown under eight different			
	medium conditions and harvested at 5 time points (18, 24,			
	48, 72, 96 hours). The poly-A sequence was trimmed off			
	before ligating to vector."			
ORIGIN				
Query Match	8.7%;	Score 88.4;	DB 7;	Length 515;
Best Local Similarity	51.0%;	Pred. No. 3.6e-16;		
Matches	209;	Conservative 0;	Mismatches 201;	Indels 0; Gaps 0;
QY	14	TCAAGTCGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAGCGGCGGTGACACA	73	
DB	58	TCAAGTTGTCGCGGTTCAAGCTGAGCCTGTCTGGAATGACCTACAAGCGGGGTTTCCA	117	
QY	74	AAACCATTTGAGTTGATGGAAGAACGACGACGTAATATGCTGCTGATCGCCTTTCCGG	133	
DB	118	AGGTCACTCTACTGATCAAGACGCTGCCAAAAACGGCGCCCAAGGTGTCGGTTTCCCG	177	
QY	134	AAACTTGATTCAGGCTACCCATGTTTCTTTGGCTTGACTCACGACATGGGCAATGC	193	

DB	178	AGGTATTCATCCCTGGTTATCCATGAGATATCTGGACTAATCCGGCTATTGAGATGTTT	237	
QY	194	AATTGTACGCCAATATCCATGAGACTCATTTGAGATTGGATGGCCCTCAAGCTAAGCGCA	253	
DB	238	CCTTCATGAACGAGTACTTCGAGAACTCCATGGAAGGAATCGAGAGATGGAACGCA	297	
QY	254	TTTCAGATGACGCCAAGCGGTTGGGATCATGCTCACCCTGGGGATGATGAACGGGTGG	313	
DB	298	TCAAAATTGCGCTGCGCGGAAGCGGGTGTATTCATCGTTCTTAGGGTATAGTAACGGTACC	357	
QY	314	GTGGCACCCTTTACATCAGTCAGTGGTTTCATAGGCGATATGTTGACACCATTTGGGGCCC	373	
DB	358	GAGATCATTTGATCATTTGCTCAGTCATTTATCGATCCTACTGGTACTATCGTCCACCACC	417	
QY	374	GGCGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTCCGCGAAGG	423	
DB	418	GACGGAAGATCAAGCCGACTCATGTGAGAGAAGCTATTGGGAGACGGG	467	
RESULT 12				
LOCUS	CF453420	655 bp	mRNA	linear EST 01-JAN-2004
DEFINITION	fts36_1_124 Fusarium verticillioides fcc1 mutant subtraction			
ACCESSION	library Gibberella moniliformis cDNA, mRNA sequence.			
VERSION	CF453420			
KEYWORDS	CF453420.1	GI:40546865		
SOURCE	EST.			
ORGANISM	Gibberella moniliformis			
	Gibberella moniliformis			
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
	Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.			
REFERENCE	1 (bases 1 to 655)			
AUTHORS	Pirttila,A.M., McIntyre,L.M., Payne,G.A. and Woloshuk,C.P.			
TITLE	Comparison of gene expression in the wild type and the fcc1 mutant			
	of Fusarium verticillioides reveals expressed sequence tags			
	associated with fumonisin biosynthesis			
JOURNAL	Unpublished (2003)			
COMMENT	Contact: Woloshuk CP			
	Department of Botany and Plant Pathology			
	Purdue University			
	915W. State Street, West Lafayette IN 47907-2054, USA			
	Tel: 765 494 3450			
	Fax: 765 494 0363			
	Email: woloshuk@purdue.edu			
	Insert Length: 655 Std Error: 0.00			
	Plate: ft row: O column: 11			
	Seq primer: T7.			
FEATURES	Location/Qualifiers			
Source	1. 655			
	/organism="Gibberella moniliformis"			
	/mol_type="mRNA"			
	/strain="FT536"			
	/db_xref="taxon:117187"			
	/clone_lib="Fusarium verticillioides fcc1 mutant			
	subtraction library"			
	/note="Vector: pGem-T-Easy; Site_1: EcoRI; Site_2: EcoRI;			
	Fungus was grown on cracked corn. RNA was isolated using			
	phenol lici method. PolyA RNA was obtained with Oligotex			
	mRNA spin columns (Qiagen). Subtracted from wild type RNA			
	with PCR select cDNA subtraction kit (Clontech),			
	amplified, cloned into pGem-T-EZ and transformed to E. coli			
	DH5 alpha cells."			
ORIGIN				
Query Match	8.6%;	Score 87;	DB 6;	Length 655;
Best Local Similarity	48.6%;	Pred. No. 1.1e-15;		
Matches	268;	Conservative 0;	Mismatches 280;	Indels 3; Gaps 1;
QY	15	CAAGTCGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAGCGGCGGTGACACA	74	
DB	3	CAAGCGCGTGTGTACCTCCGAGCCTGATGTTTGATCTTAGGGCGGTGTCGCAA	62	

[illegible]

RESULT	13
LOCUS	DR631467
DEFINITION	DR631467 699 bp mRNA linear EST 11-JUL-2005 EST1021595 FVI Gibberella moniliformis cDNA clone FVIE956, mRNA sequence.
ACCESSION	DR631467
VERSION	DR631467
KEYWORDS	GI:70706307
SOURCE	EST.
ORGANISM	Gibberella moniliformis Gibberella moniliformis Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella. 1 (bases 1 to 699)
REFERENCE	Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y., Utterback,T., Smith,S., Feldblyum,T., Glenn,A.B., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A. Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster Unpublished (2005) Contact: Brown, D.W. USDA/ARS/NCAUR
TITLE	JOURNAL
COMMENT	USDA 1815 N. University St, Peoria, IL 61604, USA Tel: 309 681 6230 Fax: 309 681 6689 Email: brownnd@ncaur.usda.gov TIGR sequence name: FVIE956TH Seq primer: AAT TAA CCC TCA CTA AAG GG.
FEATURES	Location/Qualifiers 1..699

```

FEATURES
    source
        location/Qualifiers
            1. .699
                /organism="Gibberella moniliformis"
                /mol_type="mRNA"
                /strain="m3125"
                /db_xref="taxon:117187"
                /clone="FVIE956"
                /tissue_type="mycelia"

```

/clone\_lib="Fv1"  
/note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI;  
Site\_2: XhoI; anamorph: Fusarium verticillioides. Library  
Fv1 was prepared from growth on excised maize seedling  
roots and shoots. The roots and shoots were prepared as  
described above and then inoculated by dipping briefly in  
a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated  
roots and shoots were then incubated on moistened Whatman  
#1 filter paper at room temperature under sterile  
conditions for 4 days. Fungal mycelial was collected,  
frozen in liquid nitrogen, ground to a powder, and then  
added to TRIzol Reagent (Invitrogen, Carlsbad CA) at  
approximately 1 g mycelia per 10 ml TRIzol. The cDNA was  
directionally ligated into the pBluescript II SK(+) XR  
vector (cDNA Synthesis Kit; Stratagene)."

**ORIGIN**

Query Match	8.6%	Score 87;	DB 8;	Length 699;
Best Local Similarity	48.6%	Pred. No. 1.1e-15;		
Matches 268; Conservative	0;	Mismatches 280;	Indels 3;	Gaps 1;

QY	15	CAAGTCGCGCTGCGTGCAGAGCCGCCCGCATCTACATGGAATTGGAGCGGACGGTGCACA	74
Db	68	CAAGCGCGCTGCTGTCACTCCGAGCCTGATGGTTGATCTTGAGGGCGGTGTTCGCAA	127
QY	75	AACCATGAGTTGATGGAAGAAGACGACGTAATTAATGCTCGTCTGATCGCCTTCCGGA	134
Db	128	GACCATGACTTCATCAACGAGGCCGCGCAAGCTGGCTGCAAACTCGTCCTTCCGGA	187
QY	135	AACTTGATTCAGGCTACCCATGTTGTTCTTGGCTTGACTCACCGATGGGCAATGCA	194
Db	188	AGTTTGATCCCTGGATATCCTTACTGATGTGGAAGTCACTACCTCCAATCCCTCCC	247
QY	195	ATTGTACGCCCATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCAT	254
Db	248	CATGCTGAAGCGCTACCGCGAAGAACTCCATGGCCGTGACTCTGAGAAATGCGCGTAT	307
QY	255	TTGAGATGACGCCAAGCGGTTGGAAATCATGTGTCACTGGGGATGAGTGAACGGGTGG	314
Db	308	TCGTGCGCAGCCGCGGATTAACAGATCTACGTTCCTCGGCTTCTGAGATTGACCA	367
QY	315	TGGCACCCTTTACATCAGTCAGTGTTCATAGGCGATAATGGTGACACCATTGGGCCCG	374
Db	368	CGCACTCTTTACTCTGCCCCAAGTCTCATCGGCCCCGATGGCTCGTCAACAACCA	427
QY	375	GCGAAAGTTGAAACCTACTTTGTGTGAACGTACTTTGTTCGGCGAAGG--GGATGGTTC	431
Db	428	TCGCAAGATCAAGCCAACCTCACGTTGAGAAGCTGTATACGCGGATGGCTCCGGGGATAC	487
QY	432	ATCGTAGCGGTTTTCGAGACGTCGTGTGGAAGCTGGGTGCTTATGCTGTTGGGACA	491
Db	488	CTTCATGGCGGTTAGCGAGACTGACATCGGCGGTGTGGCCACGTTAACTGCTGGAGAA	547
QY	492	CCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATTCAATTGCGGC	551
Db	548	CATGAACCCCTTCTCAAGTCTCTCAACGTTTCTGCTGGTGAGCAGGTTCAACATCGCTGC	607
QY	552	TTGGCCGAGCT 562	
Db	608	TTGGCCCGTCT 618	

RESULT	14
DR631867	
LOCUS	
DEFINITION	732 bp mRNA linear EST 11-JUL-2005 EST1021995 Fv1 Gibberella moniliformis cDNA clone FVIEE80, mRNA sequence.
ACCESSION	DR631867
VERSION	DR631867.1 GI:70706707
KEYWORDS	EST.
SOURCE	Gibberella moniliformis
ORGANISM	Gibberella moniliformis
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;





```

Db      67 CAAGCCGCTGCTGTCACCTCCGAGCCTGATGGTTGGATCTTGAGGGCGGTGTCGCA 126
QY      75 AACCATGAGTTGATGGAAGAAGCAGCAGCTAATAATGCTGCTGATGCGCTTTCCGGA 134
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     127 GACCATGACTTCATCAACGAGCGCGGCAAGCTGGCTGCAAACTGCTGCTTTCCCGA 186
QY     135 AACTTGATTCAGGCTACCCATGTTTCTTGGCTTGACTCACCAGCATGGCAATGCA 194
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     187 AGTTTGATCCCTGGATATCCTTACTGATGGGAAGTCACTACCTCAATCCCTCCC 246
QY     195 ATTTGTACGCCAATACCATGAGACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGAT 254
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     247 CATGCTGAAGCGCTACCGGAGAACTCCATGGCCGCTGACTCTGAGGAATGCGCGTAT 306
QY     255 TTCAGATGCAGCCAAGCGGTTGGGAATCATGTGTCAACCTGGGGATGATGAACGGGTCCG 314
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     307 TCGTCGGCGAGCGCCGCGATTAACAGATCTAGCTCTCCCTCGGCTTCTGAGATTGACCA 366
QY     315 TGGCACCCTTTACATCAGTCAGTGGTTCATAGCGGATAATGGTGACCAATGGGCCCG 374
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     367 CGCAACTCTTTACCTCGCCCAAGTCTCATTCGCCCGGATGGCTCGTCAACCAACCG 426
QY     375 GCGAAGTTGAAACCTTACTTTTGTGAACTTGTTCGGCGAAGGGATG--TTC 431
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     427 TCGCAAGATCAAGCAACTCAGTTGAGAACTTGTATACGGCGATGCTCCGGGATAC 486
QY     432 ATCGTAGCGGTTTTCGAGACGTCGTTGGAAGGCTGGGTGCTTATGCTGTGGAGCA 491
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     487 CTTCATGGCCGTTAGCGAGACTGACATCGGCCGTGTGGCCAGCTTAACTGTGGAGAA 546
QY     492 CCTTCACCGCTAACAATAATACGCTTTGTATGACAAATGAAGATTCAATTGTGCGGC 551
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     547 CATGAACCTCTTCTCAAGTCTCTCAAGTTTCTGTGTGAGCAGGTTCAATCGCTGC 606
QY     552 TTGGCCGAGCT 562
      ||||| |||||
Db     607 TTGGCCGCTCT 617

```

Search completed: April 27, 2006, 01:35:53  
 Job time : 4780.42 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 00:01:50 ; Search time 5473.13 Seconds  
(without alignments)  
10531.310 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1014  
Sequence: 1 atgaagaagctatcaaggt.....cgcaactcgagaatttga 1014

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_stb:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_htg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1014	100.0	1014	3	AY487562	AY487562 Unculture
2	1014	100.0	1014	6	AX189646	AX189646 Sequence
3	499.4	49.3	1014	3	AY487445	AY487445 Unculture
4	497.8	49.1	1014	3	AY487449	AY487449 Unculture
5	427.4	42.1	110000	1	CP000075_00	CP000075 Pseudomon
6	412.8	40.7	110000	1	AE016853_02	Continuation (3 of
7	399.6	39.4	999	3	AY487512	AY487512 Unculture
8	396.4	39.1	1002	3	AY487492	AY487492 Unculture
9	364.2	35.9	1038	3	AY487427	AY487427 Unculture
10	319	31.5	1011	3	AY487546	AY487546 Unculture
11	311.4	30.7	1026	3	AY487438	AY487438 Unculture
12	309.4	30.5	1077	3	AY487522	AY487522 Unculture
13	308.4	30.4	1047	3	AY487430	AY487430 Unculture
14	307	30.3	1062	3	AY487543	AY487543 Unculture
15	305.4	30.1	1041	3	AY487497	AY487497 Unculture
16	305.4	30.1	1041	6	AX189644	AX189644 Sequence
17	304.8	30.1	1065	3	AY487496	AY487496 Unculture
18	300.4	29.6	1062	3	AY487552	AY487552 Unculture

19	294	29.0	110000	1	AP006618_34	Continuation (35 o
20	289.2	28.5	1050	3	AY487432	AY487432 Unculture
21	289	28.5	1014	3	AY487464	AY487464 Unculture
22	288	28.4	1014	3	AY487429	AY487429 Unculture
23	287.8	28.4	6976	1	AY885240	AY885240 Pseudomon
24	286.2	28.2	1059	6	CQ874190	CQ874190 Sequence
25	284.6	28.1	1011	3	AY487507	AY487507 Unculture
26	283.4	27.9	1017	3	AY487558	AY487558 Unculture
27	283	27.9	1017	3	AY487475	AY487475 Unculture
28	283	27.9	1035	6	AR036234	AR036234 Sequence
29	283	27.9	1200	6	AR036235	AR036235 Sequence
30	283	27.9	1200	6	E12616	E12616 DNA encodin
31	281.4	27.8	1017	3	AY487536	AY487536 Unculture
32	281	27.7	1014	3	AY487463	AY487463 Unculture
33	280.2	27.6	1005	3	AY487559	AY487559 Unculture
34	279.6	27.6	105027	1	AJ617740	AJ617740 Pseudomon
35	278.4	27.5	1014	3	AY487515	AY487515 Unculture
36	276.6	27.3	1014	3	AY487473	AY487473 Unculture
37	274.4	27.1	984	3	AY487523	AY487523 Unculture
38	273	26.9	1017	3	AY487436	AY487436 Unculture
39	271.4	26.8	1017	3	AY487480	AY487480 Unculture
40	269.8	26.6	1017	3	AY487531	AY487531 Unculture
41	263.4	26.0	1017	3	AY487448	AY487448 Unculture
42	261	25.7	110000	1	BA000040_70	Continuation (71 o
43	258.8	25.5	1008	3	AY487545	AY487545 Unculture
44	257	25.3	1041	3	AY487481	AY487481 Unculture
45	256	25.2	1059	3	AY487532	AY487532 Unculture

ALIGNMENTS

RESULT 1	AY487562	1014 bp	DNA	linear	ENV 05-APR-2004
LOCUS	AY487562	1014 bp	DNA	linear	ENV 05-APR-2004
DEFINITION	Uncultured organism clone 2A6 nitrilase (BD5220) gene, complete cds.				
ACCESSION	AY487562				
VERSION	AY487562.1	GI:40890328			
KEYWORDS	ENV.				
SOURCE	uncultured organism				
ORGANISM	uncultured organism				
REFERENCE	1 (bases 1 to 1014)				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chl,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)				
JOURNAL	2 (bases 1 to 1014)				
REFERENCE	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA				
FEATURES	Location/Qualifiers				
source	1..1014				
gene	1..1014				
CDS	1..1014				
	/locus tag="BD5220"				
	/locus tag="BD5220"				
	/locus tag="BD5220"				
	/BC_number="3.5.5.7"				
	/codon_start=1				
	/evidence=experimental				
	/product="nitrilase"				



/protein\_id="AAR97509.1"  
/db\_xref="GI:40890329"  
/translation="MKEAIKVACVOAPIYMDLEATVDKTIELMEEAARNRARIAP  
ETWIPGYPWFMLDPSAMAMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMS  
RVGGLTLYISQWFIGDNGDTIGARKLKPTFEVERTLFEGDGSILAVFETSVGRIGLC  
CWEHLQPLTKYALYAQNEEIHCAWPSFSLYPNAKALGPDVNVAASTRIVAVEGQCFV  
LASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYANL  
DPGVRILAKMAADPAGHYSRPDITRLLIDRSPKLPVEIEGDLRPYALGKASETGAQL  
EEI"

ORIGIN

Query Match	100.0%;	Score 1014;	DB 3;	Length 1014;
Best Local Similarity	100.0%;	Pred. No. 1.1e-290;		
Matches 1014;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG	60	
Db	1	ATGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG	60	
QY	61	GCGACGGTGAACAAAACCATTTGATGATGAGAAGCAGCAGCAGTAAATGCTCGCTG	120	
Db	61	GCGACGGTGAACAAAACCATTTGATGATGAGAAGCAGCAGCAGTAAATGCTCGCTG	120	
QY	121	ATGCGCTTTCCGGAACCTTGATTTCCAGGCTACCCATGCTTCTTTGGCTTGACTCACCA	180	
Db	121	ATGCGCTTTCCGGAACCTTGATTTCCAGGCTACCCATGCTTCTTTGGCTTGACTCACCA	180	
QY	181	GCATGGGCAATGCAATTTGTATCGCCCAATACCATGAGAATCTCATTTGAGTTGATGGCCCT	240	
Db	181	GCATGGGCAATGCAATTTGTATCGCCCAATACCATGAGAATCTCATTTGAGTTGATGGCCCT	240	
QY	241	CAAAGCTAAGCGCATTTAGATGACAGCCAAAGCGTTGGGAATCATGCTCACCTGGGGATG	300	
Db	241	CAAAGCTAAGCGCATTTAGATGACAGCCAAAGCGTTGGGAATCATGCTCACCTGGGGATG	300	
QY	301	AGTGAACGGGTGCGTGACACCTTTACATCAGTCACTGCTCATAGGCGATTAATGCTGAC	360	
Db	301	AGTGAACGGGTGCGTGACACCTTTACATCAGTCACTGCTCATAGGCGATTAATGCTGAC	360	
QY	361	ACCATTTGGGGCCCGGCGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTTGGCGAA	420	
Db	361	ACCATTTGGGGCCCGGCGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTTGGCGAA	420	
QY	421	GGGGATGTTTCATCGCTAGCGGTTTTTCGAGACGTCGTTGGAAGGCTGGTGGCTTATGC	480	
Db	421	GGGGATGTTTCATCGCTAGCGGTTTTTCGAGACGTCGTTGGAAGGCTGGTGGCTTATGC	480	
QY	481	TGTTGGGAGCACCTTCAACCGCTTAACAAATACCGCTTTGTATGCACAAATGAAGATTT	540	
Db	481	TGTTGGGAGCACCTTCAACCGCTTAACAAATACCGCTTTGTATGCACAAATGAAGATTT	540	
QY	541	CATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCT	600	
Db	541	CATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCT	600	
QY	601	GATGTCAATGTAGCGGCGCTCTGAAATCTATGCCGTTGAAGGCAATGCTTCTGTAACCG	660	
Db	601	GATGTCAATGTAGCGGCGCTCTGAAATCTATGCCGTTGAAGGCAATGCTTCTGTAACCG	660	
QY	661	TGCGTGGCGCTGTTTCAAAATCCATGATGATGCTTTGTACAGATGACGAAAAAGAT	720	
Db	661	TGCGTGGCGCTGTTTCAAAATCCATGATGATGCTTTGTACAGATGACGAAAAAGAT	720	
QY	721	GCGTTGCTTCTGGCTGCTGAGCACTCACGATATAGGGCTGATGGTGTGATTTG	780	
Db	721	GCGTTGCTTCTGGCTGCTGAGCACTCACGATATAGGGCTGATGGTGTGATTTG	780	
QY	781	GTCGCGCCTCTTGCCGAAATGAAGGGTATTCTTACCGCAACCTTGATCCTGAGTA	840	
Db	781	GTCGCGCCTCTTGCCGAAATGAAGGGTATTCTTACCGCAACCTTGATCCTGAGTA	840	
QY	841	CGCATCTTGCTAAATGGCGGACACCTGCTGTCATTATTCCTGCTCCGACATTACT	900	
Db	841	CGCATCTTGCTAAATGGCGGACACCTGCTGTCATTATTCCTGCTCCGACATTACT	900	

Db	841	CGCATCTTGCTAAATGGCGGACACCTGCTGTCATTATTCCTGCTCCGACATTACT	900	
QY	901	CGCTTGCTAATAGATCGCAGCCCTTAATTAATACCAGTGTGAATGAAGGTGATCTCGT	960	
Db	901	CGCTTGCTAATAGATCGCAGCCCTTAATTAATACCAGTGTGAATGAAGGTGATCTCGT	960	
QY	961	CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA	1014	
Db	961	CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA	1014	

RESULT 2  
AX189646  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

AX189646  
Sequence 3 from Patent WO0148175.  
AX189646  
AX189646.1  
GI:15143036

unidentified  
unidentified  
unclassified sequences.

Madden, M. D. and Chaplin, J. A.  
Methods for producing enantiomerically pure g(a)-substituted  
carboxylic acids  
Patent: WO 0148175-A 3 05-JUL-2001;  
DIVERSA CORPORATION (US)  
location/Qualifiers

CDS

1..1014  
/note="unidentified DNA"  
/mol\_type="taxon:32644"  
/db\_xref="taxon:32644"  
/note="Description of Unknown Organism: Obtained from an  
environmental s amp1"  
1..1014  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC50777.1"  
/db\_xref="GI:15143037"

translation="MKEAIKVACVOAPIYMDLEATVDKTIELMEEAARNRARIAP  
ETWIPGYPWFMLDPSAMAMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMS  
RVGGLTLYISQWFIGDNGDTIGARKLKPTFEVERTLFEGDGSILAVFETSVGRIGLC  
CWEHLQPLTKYALYAQNEEIHCAWPSFSLYPNAKALGPDVNVAASTRIVAVEGQCFV  
LASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYANL  
DPGVRILAKMAADPAGHYSRPDITRLLIDRSPKLPVEIEGDLRPYALGKASETGAQL  
EEI"

ORIGIN

Query Match	100.0%;	Score 1014;	DB 6;	Length 1014;
Best Local Similarity	100.0%;	Pred. No. 1.1e-290;		
Matches 1014;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG	60	
Db	1	ATGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG	60	
QY	61	GCGACGGTGAACAAAACCATTTGATGATGAGAAGCAGCAGCAGTAAATGCTCGCTG	120	
Db	61	GCGACGGTGAACAAAACCATTTGATGATGAGAAGCAGCAGCAGTAAATGCTCGCTG	120	
QY	121	ATGCGCTTTCCGGAACCTTGATTTCCAGGCTACCCATGCTTCTTTGGCTTGACTCACCA	180	
Db	121	ATGCGCTTTCCGGAACCTTGATTTCCAGGCTACCCATGCTTCTTTGGCTTGACTCACCA	180	
QY	181	GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT	240	
Db	181	GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT	240	
QY	241	CAAAGCTAAGCGCATTTAGATGACAGCCAAAGCGTTGGGAATCATGTCACCCCTGGGATG	300	
Db	241	CAAAGCTAAGCGCATTTAGATGACAGCCAAAGCGTTGGGAATCATGTCACCCCTGGGATG	300	
QY	301	AGTGAACGGGTGCGTGACACCTTTACATCAGTCACTGCTCATAGGCGATTAATGCTGAC	360	

Db 301 AGTGAACGGGTGCGTGAGACCCCTTACATCAGTCAGTGTTCATAGCGCATATGSGTAC 360  
QY 361 ACCATTGGGGCCCGCGAAGTGAACCTACTTTGTGTAACGTACTTTGTTCGGCGAA 420  
Db 361 ACCATTGGGGCCCGCGAAGTGAACCTACTTTGTGTAACGTACTTTGTTCGGCGAA 420  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGGTGCTTATGC 480  
Db 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGGTGCTTATGC 480  
QY 481 TGTGGGAGCACCCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGATT 540  
Db 481 TGTGGGAGCACCCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGATT 540  
QY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGGAAAGCCCTGGGCCCT 600  
Db 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGGAAAGCCCTGGGCCCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTACTAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCGTACTAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGCATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCCATGATGCATATGCTTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTGTCTTGTGCTGTGTGTGACACTACGATCATATAGGGCTGATGCTGTGACTTG 780  
Db 721 GCGTGTCTTGTGCTGTGTGTGACACTACGATCATATAGGGCTGATGCTGTGACTTG 780  
QY 781 GTGCGGCTCTTGCCGAAATGAAGAGGATTTCTTACGCCAAACCTTGATCTCGAGTA 840  
Db 781 GTGCGGCTCTTGCCGAAATGAAGAGGATTTCTTACGCCAAACCTTGATCTCGAGTA 840  
QY 841 CGCATCTTGTCTAAATGGCGGACAGCCCTGTGTCTATTATCCCGTCCCGACATTACT 900  
Db 841 CGCATCTTGTCTAAATGGCGGACAGCCCTGTGTCTATTATCCCGTCCCGACATTACT 900  
QY 901 CGCTGTCTAATAGATCGACAGCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTGT 960  
Db 901 CGCTGTCTAATAGATCGACAGCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTGT 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGTA 1014  
Db 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGTA 1014  
  
RESULT 3  
AY487445 1014 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A8 nitrilase (BD5274) gene, complete  
DEFINITION cds.  
ACCESSION AY487445  
VERSION AY487445.1 GI:40890094  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
AUTHORS 1 (bases 1 to 1014)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,B., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955

FEATURES  
Source Directors Place, San Diego, CA 92121, USA  
1..1014  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A8"  
/environmental\_sample  
1..1014  
/locus\_tag="BD5274"  
1..1014  
/locus\_tag="BD5274"  
/locus\_tag="BD5274"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/transl\_except=(pos:1..3,aa:Met)  
/product="nitrilase"  
/protein\_id="AAR97392.1"  
/db\_xref="GI:40890095"  
/translation="MKEAIKVACVQAAPVFLDLDATVDKTVALLIEAARNGARLIAFP  
ETWIDGYFWFLMIDSPAWGMOFVRHYHNSLVDSPOAKRISBAQAGIYVALGYSE  
RVSGTLVWGQWLIDDKGETAGLRKLPKTHVERTLPFGDGSLSLFTDPLVGLGLC  
CWEHLQPLSKYALYAQNEIHFAPWPSFSIYRQATEVLEPEVNVVAASRIYAVEGQCFV  
LASCALVSPMEIEMLCCTDESKSLDQAGGYSRIIGPDGSLARPGENEBGILYATL  
DPAARIYAKTAADPAGHYSRPDYTRLLINRSANQPVVEVGREIPASAQFEVEAPGY  
EGD"  
  
ORIGIN  
Query Match 49.3%; Score 499.4; DB 3; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 3.6e-137;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;  
  
QY 2 TGAAGAAGCTATCAAGGTGCGCTGCTGCAAGCCCGCATCTACATGATTTGAGG 61  
Db 2 TGAAGAAGCAATCAAAAGTAGCTGTGTGCAAGCAGCTCCAGTCTTTCTGCACTGAGC 61  
QY 62 CGACGTGACAAACCATTTGATGATGGAAGAAGCAGCAGCACTAATTAATGCTGCTGA 121  
Db 62 CCACAGTGACAAAGACCGTGCCTGATGAGAGGACAGCCGTAACGGCGACAGCCTAA 121  
QY 122 TCGCCTTCCGAAACTTGATTCAGGCTAACCCATGTTCTTTGGCTTGACTCACCAG 181  
Db 122 TCGCCTTCCAGAGACCTGATTCAGGCTAACCCATGTTCTTTGGCTTGACTCACCAG 181  
QY 182 CATGGCAATGCAATTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTC 241  
Db 182 CTTGGGGATGCAATTGTGCGCCGATACAGAGAACTCACTGGTCTCTGACAGCCCTC 241  
QY 242 AAGCTAAGCGATTTACAGATGACGCCAGCGGTTGGAAATCATGTGACCCCTGGGATGA 301  
Db 242 AGGCCAAGCGATCAGTGAGCGCGCCGCGCGGTATATACGTGCGCTAGGGTACA 301  
QY 302 GTGAACGGGTGCTGGACACCTTTACATCAGTGTGTTCATAGCGGATAATGTGTGACA 361  
Db 302 GCGAACGGGTGAGCGGAACCTCTACATGGGGCAGTGGCTCATTTGAGCATTAAGGGCGAAA 361  
QY 362 CCATTGGGGCCCGCGAAGTGAACCTACTTTGTGTAACGTACTTTGTTCGGCGAAG 421  
Db 362 CAGCTGGGCTGCGCGGAAAGCTGAACCAACCATGTAGAGCGGAACCTCTTCGGTGAAG 421  
QY 422 GGGATGTTTCATCGCTAGCGGTTTTCAGACGTCGTGTGGAAGGCTGGGCTTATGCT 481  
Db 422 GCGACGATCATCCCTTTCCACTTTGACACACCGTTGGGGGTGCTGGGCGGACTCTGCT 481  
QY 482 GTTGGAGCACCCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGAGATT 541  
Db 482 GTTGGAGCACCCTTCAACCGCTTTTGAATATGCGCTCTACGACAGAAAGAAATAC 541  
QY 542 ATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCTTAATGCGCGGAAAGCCCTGGGCTG 601  
Db 542 ACTTGCAGGCTTGCGCTAGCTTACGATCTACCGTCAAGCGACAGAAAGTCTTGACACAG 601  
QY 602 ATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGCGAATGCTTGTACTAGCGT 661

Db 602 AAGTAAATGCGAGCTTCTCGATCTACGCCGTGGAAGGCGAGTGTGTTGTTCTCGCTT 661  
QY 662 CGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCATG 721  
Db 662 CCGCGCGCTCGTCTCGCCAGAGATGATGAAATGCTCTGCACTGACGAAAAGCAAGCACA 721  
QY 722 CGTTGCTTCTGCTGTGTGTGACACATCAGTATCATAGGGCCCTGATGTGTGACTTGG 781  
Db 722 GCCTTCTTCAAGCCGGCGGGGTACTCCGCATTAATCGTCCGATGGCAGCAGACTAG 781  
QY 782 TCGCGCCTTTCGCGAAATGAGAGGGTATCTCTACGCAAACTTGATCTTGAGTAC 841  
Db 782 CGCGCCCTTGGGCGAAAACGAGAGAGTATCTCTATGCGCACTTGGAACCTGCGCTC 841  
QY 842 GCATCCTTGCTAAATGGCGGAGACCTGTGTCATTATTCGCCGTCGACATTACTC 901  
Db 842 GAATCTATGCAAGACCGCAGCTGATCCAGCGGGCACTACTCCAGACGAGCGTCACTC 901  
QY 902 GCTTGCTAATAGATCGCAGCCCTAATTAACCGGTAGTTGAATTGAAGTGATCTTGTG 961  
Db 902 GGCTGCTGATCAATCGCAGTGCCCAATCAGCCAGTCGTAGAGTTGGAAGGAAATACCTG 961  
QY 962 CTTAAGCTTTGGGTAAGCGCTCTGAGACGGGTGCGCACTCGAAGAAATTGA 1014  
Db 962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCGCCCGGGTACGAAGCGCATTTGA 1014

RESULT 4  
AY487449 1014 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A7 nitrilase (BD5275) gene, complete  
DEFINITION cds.  
ACCESSION AY487449  
VERSION AY487449.1 GI:40890102  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
unclassified; environmental samples.  
REFERENCE 1 (bases 1 to 1014)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1014)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
REFERENCE Directors Place, San Diego, CA 92121, USA  
AUTHORS location/Qualifiers  
FEATURES  
source 1..1014  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A7"  
/environmental\_sample  
1..1014  
/locus\_tag="BD5275"  
1..1014  
/locus\_tag="BD5275"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/transl\_except=(pos:1, .3,aa:Met)  
/product="nitrilase"  
/protein\_id="AAR97396.1"  
/db\_xref="GI:40890103"  
/translation="MKEAIKVAQVQAPVFLDLDAIVDKTVALIEEAPRNGARLIAPF

ETWIPGYPMFLMLDSPAWMQFVRRYHENSILVDSPOAKRISSEAQRAGIYVALGYSE  
RVSGTLYMGOMLIDDKETAGLRKRLKPTHERVTLFEGDGSSLSFTDPLGLGLC  
CWEHLQPLSKYLAQNEBIHFAAPSPSIYROATEBLGPEVNVYASRIYAVEGQCFV  
LASCALVSPENIEMLCIDESKSHSLQAGGYSRIIGPDGDLARPLGENEGLIYATL  
DPAARIYAKTAADPAHYSRDPVTRLLINRSANQPVVEVGREIPASQGFVEEAAPGY  
GGD"

ORIGIN

Query Match 49.1%; Score 497.8; DB 3; Length 1014;  
Best Local Similarity 68.2%; Pred. No. 1.1e-136;  
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGGTGCGCTGCGTCAAGCCGCCATCTACATGATTTGGAGG 61  
Db 2 TGAAGAAGCAATCAAGTACCTGTGTGCAAGCAGCTCCAGTCTTCTCGACCTGGACG 61  
QY 62 CGACGGTGAACAAAACCAATGAGTTGATGGAAGAAGCAGCAGCTAATATGCTCGTGA 121  
Db 62 CCACAGTGAACAAGACCGTCCGCTGATTGAGAGGACAGCCGTTAACGGCGCACGCTTAA 121  
QY 122 TCGCCTTCCGAAACTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACAG 181  
Db 122 TCGCCTTCCAGAGACCTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACAG 181  
QY 182 CATGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241  
Db 182 CTTGGGGATGCAATTCTGTGCGCGATACCAAGAACTCACTGTCTCGACAGCCCTC 241  
QY 242 AAGCTAAGCGCATTTTCAGATGACGCCAAGCGGTTGGAAATCATGTCAACCTGGGATGA 301  
Db 242 AGGCCAAGCGCATCAGTAGAGCCGCCAGCGCGGTATATAGTCGCGCTAGGGTACA 301  
QY 302 GTGAACGGGTGCGGTGCGACCCCTTACATCAGTCACTGTTGATGAGGATATAGGTGACA 361  
Db 302 GCGAAGCGGTGAGCGGAACCTCTTACATGGGGCAGTGGCTCATTTGACGATAAGGCGAAA 361  
QY 362 CCATTGGGGCGCGCGGAAGTTGAACCTACTTTGTGTAACGTACTTTGTCGGCGAAG 421  
Db 362 CAGCTGGGCTGCGCGGAAAGCTGAACCAACCATGTAGAGCGAACCTCTTCGGTGAAG 421  
QY 422 GGGATGTTCACTGCTAGCGGTTTTCAGACGCTGTGTGGAAGCGTGGGCTATGCT 481  
Db 422 GCGACGATATCCCTTTCACATTTCGACACACCGTTGGGGGTGCTGGCGGACTGTGCT 481  
QY 482 GTTGGAGACCTTCAACCGCTAACAAAATACGCTTTGTATGACAAAATGAAGAGATTG 541  
Db 482 GTTGGAAACACTTACAACCTCTTTCGAATATGCGCTCTACGACAGAACGAGAAATAC 541  
QY 542 ATTGTGCGGCTTGCGCGAGCTTTAGCCTTATCTTAATGCGCGAAAGCCCTGGGCGCTG 601  
Db 542 ACTTGCAGGCTGCGCTAGCTTCAAGCATCAACGTCAGACAGAAAGTCTTGGACACAG 601  
QY 602 ATGTCAATGATGCGGCTCTCGAATCTATGCGGTTGAAGGCGCAATGCTTCTACTAGCGT 661  
Db 602 AAGTAAATGTCGACGCTTCTCGATCTACGCGGTGAAGGCGAGTGTGTTGTTCTCGCTT 661  
QY 662 CGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCATG 721  
Db 662 CCGTGGCGCTCGTCTGCGCAAGATGATGCAATATGCTCTGCACTGACGAAAGCAAGCACA 721  
QY 722 CGTTGCTTCTGCGTGTGTGACACTCAAGTATCATAGGGCGCTGATGTGTGACTTGG 781  
Db 722 GCCTTCTTCAAGCCGGCGGGGTACTCCGCAATTAATGCGGTCCGATGGCAGGACCTAG 781  
QY 782 TCGCGCCTCTTGCAGAAATGAGAGGGTATTTCTTACGCAAACTTGATCTGAGTAC 841  
Db 782 CGCGCCCTTGGCGAAAACGAGAGAGTATTTCTTATGCGCACTTGACCCCTGCGCTC 841  
QY 842 GCATCCTTGCTAAATGGCGGAGACCCCTGTGTCATTATTCGCCGTCGACATTACTC 901  
Db 842 GAATCTATGCAAGACCGCAGCTGATCCAGCGGGCACTACTCCAGACGAGCGTCACTC 901  
QY 902 GCTTGCTAATAGATCGAGCCCTAAATTACCGGTAGTTGAATTGAAGTGATCTTGTG 961



Db 902 GGCTGTCATCATCGCAGTCCCAATCAGCCAGTCGTAGAGTTGACGGGAATACCTG 961  
QY 962 CTTACGCTTTGGGTAAGCGCTGTAGACGGGTGCGCAACTCGAAGAATTGA 1014  
Db 962 CATCGGCCCAAGGCTTTGAAGTTGAGCGCGCCCCCGGCTACGAGGCGATTGA 1014

RESULT 5  
CP000075\_00/c  
WPCOMMENT

Sequence split into 61 fragments LOCUS CP000075 Accession CP000075

Fragment Name	Begin	End
CP000075_00	1	110000
CP000075_01	100001	210000
CP000075_02	200001	310000
CP000075_03	300001	410000
CP000075_04	400001	510000
CP000075_05	500001	610000
CP000075_06	600001	710000
CP000075_07	700001	810000
CP000075_08	800001	910000
CP000075_09	900001	1010000
CP000075_10	1000001	1110000
CP000075_11	1100001	1210000
CP000075_12	1200001	1310000
CP000075_13	1300001	1410000
CP000075_14	1400001	1510000
CP000075_15	1500001	1610000
CP000075_16	1600001	1710000
CP000075_17	1700001	1810000
CP000075_18	1800001	1910000
CP000075_19	1900001	2010000
CP000075_20	2000001	2110000
CP000075_21	2100001	2210000
CP000075_22	2200001	2310000
CP000075_23	2300001	2410000
CP000075_24	2400001	2510000
CP000075_25	2500001	2610000
CP000075_26	2600001	2710000
CP000075_27	2700001	2810000
CP000075_28	2800001	2910000
CP000075_29	2900001	3010000
CP000075_30	3000001	3110000
CP000075_31	3100001	3210000
CP000075_32	3200001	3310000
CP000075_33	3300001	3410000
CP000075_34	3400001	3510000
CP000075_35	3500001	3610000
CP000075_36	3600001	3710000
CP000075_37	3700001	3810000
CP000075_38	3800001	3910000
CP000075_39	3900001	4010000
CP000075_40	4000001	4110000
CP000075_41	4100001	4210000
CP000075_42	4200001	4310000
CP000075_43	4300001	4410000
CP000075_44	4400001	4510000
CP000075_45	4500001	4610000
CP000075_46	4600001	4710000
CP000075_47	4700001	4810000
CP000075_48	4800001	4910000
CP000075_49	4900001	5010000
CP000075_50	5000001	5110000
CP000075_51	5100001	5210000
CP000075_52	5200001	5310000
CP000075_53	5300001	5410000
CP000075_54	5400001	5510000
CP000075_55	5500001	5610000
CP000075_56	5600001	5710000
CP000075_57	5700001	5810000
CP000075_58	5800001	5910000
CP000075_59	5900001	6010000

LOCUS CP000075.60 6000001 6093698  
DEFINITION CP000075 Pseudomonas syringae pv. syringae B728a, complete genome.  
ACCESSION CP000075 AABP02000000 AABP02000001-AABP02000026  
VERSION CP000075.1 GI:63253978  
KEYWORDS  
SOURCE Pseudomonas syringae pv. syringae B728a  
ORGANISM Pseudomonas syringae pv. syringae B728a  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.  
REFERENCE  
1 (bases 1 to 6093698)  
Feil,H., Feil,W.S., Chain,P., Larimer,F., Dibartolo,G., Copeland,A., Lykidis,A., Trong,S., Nolan,M., Goldsman,E., Thiel,J., Malfatti,S., Loper,J.E., Lapidus,A., Deter,J.C., Land,M., Richardson,P.M., Kyriides,N.C., Ivanova,N. and Lindow,S.E.  
DOE Joint Genome Institute  
Comparison of the complete genome sequences of Pseudomonas syringae pv. syringae B728a and pv. tomato DC3000  
Proc. Natl. Acad. Sci. U.S.A. 102 (31), 11064-11069 (2005)  
16043691  
2 (bases 1 to 6093698)  
Loper,J.  
Direct Submission  
Submitted (04-MAY-2005) Oregon State University, OR 97331, USA  
3 (bases 1 to 6093698)  
Feil,H., Feil,W.S. and Lindow,S.E.  
Direct Submission  
Submitted (05-MAY-2005) Departments of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA  
On or before May 12, 2005 this sequence version replaced  
gi:28876540, gi:28876539, gi:28876538, gi:28876537, gi:28876536, gi:28876535, gi:28876534, gi:28876533, gi:28876532, gi:28876531, gi:28876530, gi:28876529, gi:28876528, gi:28876527, gi:28876526, gi:28876525, gi:28876524, gi:28876523, gi:28876522, gi:28876521, gi:28876520, gi:28876519, gi:28876518, gi:28876517, gi:28876516, gi:28876515.  
URL -- http://www.jgi.doe.gov  
Contacts: Paul Richardson (microbes@cuba.jgi-psf.org)  
Steven E. Lindow (lcelab@socrates.berkeley.edu)  
Finished microbial genomes have been curated to close all gaps with greater than 98% coverage of at least two independent clones. Each base pair has a minimum q (quality) value of 30 and the total error rate is less than one per 50,000.  
FEATURES  
source  
1. .6093698  
/organism="Pseudomonas syringae pv. syringae B728a"  
/mol\_type="genomic DNA"  
/strain="B728a"  
/db\_xref="taxon:205918"  
/note="pathovar: syringae"  
1. .1536  
/locus\_tag="Psysr\_0001"  
1. .1536  
/locus\_tag="Psysr\_0001"  
/locus\_start=1  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="chromosomal replication initiator protein, DnaA"  
/protein\_id="AA135075.1"  
/db\_xref="GI:63253979"  
/db\_xref="InterPro:IPR001957"  
/db\_xref="InterPro:IPR003593"  
/translation="MSVELMQCVELLRDELPAQQNTWIRPLQVEAGDELRYAPN RFVLDWVNEKYLGRLLLELGEHGQMAPALSLIGSKRSSAPRAPAPLAASQAL SANSVSSVAPAPATAPAAVAATPAPVQNVATHDEPSRDSFDPMGASSQAARAE QRTVQVEGALKHTSYLNRTFTFENFVEGSKNQLARAAQVADNPRGYNPFLFYGV GLGKTHLMAVGNHLIKKNPNKVVYLSERFVADMKALQNAINEFKRFYRSVDAL LIDDIQFPAKERSQBEFFHTFNALLEGQOVIITSDRYPKIEGLEERLKSFRGWL TVAVEPPELETRVAIIMKKADQKVDLPHDAFFIAQRIRSNVRELGALKRVTAHSH FMGRDITIELIRSLKDLALQDKLVSDNIQRTVAEYKIKISDLKSRRSVARP ROVAMALSKELTNHSLPEIGDVFQGRDHTTVLHACRKINELKESDADIREDYKNLLRT LTT"  
1575. .2678  
gene



CDS
/locus\_tag="Psysr\_0002"
1575..2678
/locus\_tag="Psysr\_0002"
/EC\_number="2.7.7.7"
/codon\_start=1
/evidence=not\_experimental
/transl\_table=11
/product="DNA polymerase III, beta chain"
/protein\_id="AA035076.1"
/db\_xref="GI:63253980"
/db\_xref="InterPro:IPR001001"
/translacion="MHFTIQREALKPLQLVAGVERRQTLPVLSNVLLVEGQQLSL
TGTDLEVLVGRVQLIEEPAEPEGLTVPAKLMIDICKSLNDALIDIKLIDSKLIVKAG
RSRFTLSTLPANDPFTVEEGPSLFTTLVQSKLRLLIERTSFMAAOQDVRYYLNGMLI
EVSAGILRAVATDGHRLAMCSMSADIEMADRHQIVIPRKGLEMARLLTEODGTVSIV
LGQHIRATTBGEFTFTSKLVGDKPDYERVLPGGDKLVLGROALREAFSRTAILSN
EKYRGIRLQLASGQLKIQANNPEOEAEIEISVEYNGDSLIEGFNVSYLLDVLGVMTT
EOVRLIIDSNSSALVQESDNDSDAYVVMRL"
2701..3804
/locus\_tag="Psysr\_0003"
2701..3804
/locus\_tag="Psysr\_0003"
/codon\_start=1
/evidence=not\_experimental
/transl\_table=11
/product="RecF protein"
/protein\_id="AA035077.1"
/db\_xref="GI:63253981"
/db\_xref="InterPro:IPR001238"
/db\_xref="InterPro:IPR003395"
/translacion="MSLSRVSVTVGNLHPVTLSPSPRINILYGANGSGKTSVLEAII
LLGIARFRRSRLLPVIOYEQPSCTVFGQVDLAQGGHNSLVGRDRQGEFQIRIDGON
ARSAQIAELPLQLINPDSFRLTEGAPKIRQLDWGFHVEPRFMATWQRLQKALK
QNSWMLHGTLDASQAAMDRELCSAIDEDEFRAVYIKALKPVEQTLSELVELEG
TLSYRGWDKEKELSTVLASSIHRDQMGHTQAGPORADLRRLGANAADILSRGQQ
KLVLVCLRIAQGHVSVQVRGQCTIYVDDLPELSDNHRALCRLLLELRQVFITCV
DOEFLREGWQETETPVALFHVEQGRITQTHDRE"
3809..6226
/locus\_tag="Psysr\_0004"
3809..6226
/locus\_tag="Psysr\_0004"
/EC\_number="5.99.1.3"
/codon\_start=1
/evidence=not\_experimental
/transl\_table=11
/product="DNA gyrase, B subunit"
/protein\_id="AA035078.1"
/db\_xref="GI:63253982"
/db\_xref="InterPro:IPR000565;IPR001241;IPR002288;IPR003594;IPR006171;IPR01155
7;IPR011558"
/translacion="MSENQTYDSSIKVLKGLDAVRKRPGRMYIGTDDGSLHMFVE
VVDNSIDEALAGHDDISIIHPDESTIVRDNGRGI PVDVHKEEGVSAAEVIMTVLHA
GKFDNSYKVSGLHGVSVVNALSELILLTVRSKIMWECTYHGVPOEPMKIVG
ESDSTGTQIHFKPSAETPKNIHFSWDLAKRIRELSFLNSGVGIYIKDERSGKEELFK
YEGGLRAFEYLTNTKTPVNEVFHFNIQRDDIGVEIALQWNSFNENILCFNNIPQ
RDGTHLVGFRSALTRNLNNYIEQGLAKHKVATTGDDAREGLTAISVKVPDPKFS
SQTKDVLVSEVKTAVEQEMGKYFSDFLIENPEAKAVGKMDAARAREAKAREM
TRKGLDLIAGLPGKLADQEKDPALSELVVEGDSAGSAGSAKQGRNRTOAILPLKKG
ILNVEKARPDKMISQEVGTLITALGCGIGREEYNIDKLRHNIITMTDADVDGSHIR
TLTLTFPFRQPLBELIERGYIYAOPLYVKKKQEOYIKDDEAMEEYMTQSALEDAS
LHLSBAPGISGTALIEKLVNDFRMVMKTLKRLSRLYPOELTEHFVYLPITLQQLSDH
EGMQAWLALFDARLRTGEKSGLVYKASLRDRERNVMLPEVELISHGLSNVTFENRDF
FGSNDYKTVTALGAQISTLLEGAAYVRGERKKPVNEFKALAWLMAESTKRHTTQRY

Query Match 42.1%; Score 427.4; DB 1; Length 110000;
Best Local Similarity 65.9%; Pred. No. 1.3e-115;
Matches 620; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 1 ATGAAAGAAGCTATCAAGTCGCTGCGTGAAGCCGCCGATCTACATGATTGGAG 60
DB 9243 ATGAAAGAAGCTTGAAGTCGCTGTGTGACAGCTGGCCCGCTGTCTGTGATCTGGAC 9184

QY 61 GCGACGGTGCACAACCATTTGATGGAAGAAGCAGACAGTAATAATGCTGCTG 120
DB 9183 GCCACCGTAGACAAGACCATAACCTGATGAGCAGCGCGCAGCCGAGCCGCTT 9124
QY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCATGCTTCTTGCTGACTCACCA 180
DB 9123 ATTGCATTCGCCGAAACGTGATTCGGGCTATCCGTGTTTGTGGCTCGACGACCG 9064
QY 181 GCATGGGCAATGCAATTTGTACGCCCAATACCATGAGAATCATTTGAGTTGGATGGCCCT 240
DB 9063 GCCTGGAACATGCCGTTGTTGTCAGCGCTTATCACCAAGCAGTCTTGCTGACAGTGT 9004
QY 241 CAAGCTAAGCGCATTCAGATGCAGCCAGCGGTTGGGAATCATGATCAACCTGGGATG 300
DB 9003 CAGGCCGACGTATCAGCGATGCGGCTCGCCATCTCGGTCTTATGTGTGCTCGGTTAC 8944
QY 301 AGTGAACGGGTGCGTGGCACCCTTTACATCAGTCAGTGTGTTCATAGGCGATAATGTGTAC 360
DB 8943 AGCGACGCAACAAAGCGTCGCTGTATATCGGTCAGTGATCATGACGATCACGGTGAA 8884
QY 361 ACCATTGGGCGCCGCGCAAGTTGAACCTACTTTTGTGAACGTACTTGTGGCGAA 420
DB 8883 ACGTGTGGCGTGCCTCGCAAGCTCAAGGCTACTCATGTGAACGTACGATGTTGGCGAA 8824
QY 421 GGGATGTTTCATCGCTAGCGGTTTCGAGACGTCTGTTGAAGGCTGGGTGCTTATGC 480
DB 8823 GCGACGGGGCATCATTCGCTACCTTTGAACCCCTGTCGGCGTCTCGGTGCGCTTGC 8764
QY 481 TGTGGAGACACTTCAACCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540
DB 8763 TGCTGGAGCACTTCAACCGCTTCCAAAGTACGCCATGTACGACAGAACGAACAGATC 8704
QY 541 CATTTGCGGCTTGCCCGAGCTTTAGCCTTATCCTAATGCGGCGCAAGCCCTGGGCT 600
DB 8703 CACGTAGCGGCTGGCCAAAGCTTCAGCTGTACCGCAACGCCACGAGTGCCCTGGGCCCC 8644
QY 601 GATGTCAATGACGGGCTCTGAACTATAGCCGTGAAGGGCAATGCTTGTACTAGCG 660
DB 8643 GAGTCAATGACGGGCTTCCGGGTTTAGCCGACAGAAAGGCAATGCTTGTCTCGCA 8584
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATGCTTGTGTACAGATGACGAAAGCAT 720
DB 8583 CCCTGTCAATCGTTTCCCTGAAATGATGAAATGCTTGTGCAACAGTATGCCAAAGCG 8524
QY 721 GCGTGTCTTGGCTGTGTGTGACACATCATAGAGGCTGATGGTGTGACTTG 780
DB 8523 AGCCTGTGCAAGCGGCTGGCGGACGACGAGGATTTTCGTCCGAGCGAAGTATCTG 8464
QY 781 GTCGCGCTCTTGCAGAAATGAAGAGGATTTCTTACGCAAACTTGAATCCTGAGTA 840
DB 8463 GCGACCCCATTTGGCGAATGATGAGGAGGTCTGCTTACGCAACACTTGAATCCAGCGCA 8404
QY 841 CGCATCCTTGTAAATGGCGGACAGCCCTGTGTCATTATTCCTGCGACATTAAT 900
DB 8403 TTGACCTTGGCCAAAGTACGAGCCGACCCGCTGGAATTAATCACTGACGTCGCACTCACT 8344
QY 901 CGCTTGTAAATGATTCGACGACCCCTAAATTAACCGGTAGTTGA 941
DB 8343 CGTGTGATGTTCAATCCCAATCCCACTTCATGTGTATGCA 8303

RESULT 6
AE016853\_02
WPCOMMENT

Sequence split into 64 fragments	LOCUS	AE016853	Accession	AE016853
Fragment Name	Begin	End		
AE016853_00	1	110000		
AE016853_01	100001	210000		
AE016853_02	200001	310000		
AE016853_03	300001	410000		
AE016853_04	400001	510000		
AE016853_05	500001	610000		
AE016853_06	600001	710000		

AE016853_07	700001	810000
AE016853_08	800001	910000
AE016853_09	900001	1010000
AE016853_10	1000001	1110000
AE016853_11	1100001	1210000
AE016853_12	1200001	1310000
AE016853_13	1300001	1410000
AE016853_14	1400001	1510000
AE016853_15	1500001	1610000
AE016853_16	1600001	1710000
AE016853_17	1700001	1810000
AE016853_18	1800001	1910000
AE016853_19	1900001	2010000
AE016853_20	2000001	2110000
AE016853_21	2100001	2210000
AE016853_22	2200001	2310000
AE016853_23	2300001	2410000
AE016853_24	2400001	2510000
AE016853_25	2500001	2610000
AE016853_26	2600001	2710000
AE016853_27	2700001	2810000
AE016853_28	2800001	2910000
AE016853_29	2900001	3010000
AE016853_30	3000001	3110000
AE016853_31	3100001	3210000
AE016853_32	3200001	3310000
AE016853_33	3300001	3410000
AE016853_34	3400001	3510000
AE016853_35	3500001	3610000
AE016853_36	3600001	3710000
AE016853_37	3700001	3810000
AE016853_38	3800001	3910000
AE016853_39	3900001	4010000
AE016853_40	4000001	4110000
AE016853_41	4100001	4210000
AE016853_42	4200001	4310000
AE016853_43	4300001	4410000
AE016853_44	4400001	4510000
AE016853_45	4500001	4610000
AE016853_46	4600001	4710000
AE016853_47	4700001	4810000
AE016853_48	4800001	4910000
AE016853_49	4900001	5010000
AE016853_50	5000001	5110000
AE016853_51	5100001	5210000
AE016853_52	5200001	5310000
AE016853_53	5300001	5410000
AE016853_54	5400001	5510000
AE016853_55	5500001	5610000
AE016853_56	5600001	5710000
AE016853_57	5700001	5810000
AE016853_58	5800001	5910000
AE016853_59	5900001	6010000
AE016853_60	6000001	6110000
AE016853_61	6100001	6210000
AE016853_62	6200001	6310000
AE016853_63	6300001	6397126

Continuation (3 of 64) of AE016853 from base 200001 (AE016853 *Pseudomonas syringae* pv. t

Query Match 40.7%; Score 412.8; DB 1; Length 110000;  
Best Local Similarity 64.8%; Pred. No. 2.9e-111;  
Matches 612; Conservative 0; Mismatches 332; Indels 0; Gaps 0;

QY	1	ATGAAGAAGCTATCAAGTCGCTGCGTGCAGCGCCCGCATCTACATGATTTGGAG	60
DB	12501	ATGAAGCGCCACTGAAGTTGCTGCTACAGCGCCGCCCATGTTCCTGAATCTGAC	12560
QY	61	GCGACGGTGGACAAACCATTGATGTGAGAGCAGCAGTAATATGCTGCTG	120
DB	12561	GCCACGGTAGACAAAGCATTAGCCCTGATCGAGCAGGCGCGCCGAGCGGCTG	12620
QY	121	ATGCGCTTTCGGAACCTTGATTCAGGCTACCCATGGTTTCTTTGGCTTGACTCACA	180

DB	12621	ATTGATTCCCCGAGACGTGATTCGGGGCTAACCCATGGTTTGTGGCTGACGACCG	12680
QY	181	GCATGGCAATGCAATTTGTACGCCAATACATGAACTCATTGGATTGGCCCT	240
DB	12681	GCTTGAAACATGCCGTGTGTGATCGCTATACCAAGCAGTCCTGTGTGTGACAGTGCT	12740
QY	241	CAAGTAAGCATTTCAGATGACGCCAAGCGTTGGAAATCATGTACCTGGGATG	300
DB	12741	CAGCAGCAGCAGTATGATGACAGTCGTGACACAGGTATCTACGTGTGTGGTTAC	12800
QY	301	AGTGAACGGTGGTGCAACCTTTACATCACTAGTGTTCATAGGCGATAATGTGAC	360
DB	12801	AGCGAGCGCAGCAGGCTTCGCTGATATCGGTCAAGTATCATGACATCAAGCGAG	12860
QY	361	ACCATTTGGGGCCCGCGAAAGTTGAAACCTTTTGTGTAACGTACTTGTTCGGCGAA	420
DB	12861	ACCGTCGCACGCGCGCAAGCTCAAGGCGCAGCATGTGAAACGTACTTGTTCGGCGAA	12920
QY	421	GGGATGTTTCAATCGCTAGCGGTTTTCGAGACGTCTGTGAAAGCGTGGGCTTATGC	480
DB	12921	GGAGACGGGGCGTCAATGCGACCTTTGAAACACTGTAGCGTACTGGTGCCCTTTC	12980
QY	481	TGTTGGAGCACCCTTCAACCGCTAACAAATAACGCTTGTATGACAAATGAGAGATT	540
DB	12981	TGCTGGAGCACCCTGCAACCGCTCTCGAAATACCGCATGTACGCGCAGAACACAGATT	13040
QY	541	CATTGTGCGGCTTGCGCAGCTTTAGCCTTATCTAATGCGCGAAAGCCTGGGGCCT	600
DB	13041	CACGTAGCCGCTTGCGCCACGCTTACGCTGTACAGAACGCCACCGCGGCTTGGCCCA	13100
QY	601	GATGTCATGTAGCGGCTCTCGAATCTATGCGGTGAAAGGCAATGCTTCTGTAAGCG	660
DB	13101	GAGGTCAACACCGCGCTTACGCGTTATGCGCAGAAAGGCGAGTGTCTGTCAATTGCA	13160
QY	661	TCGTGCGCTCGTTTCACAATCCATGATGATATGCTTGTACAGATGACGAAAGCAT	720
DB	13161	CCGTGCGAGTGTCTCTCCGAAATGATCGATGCTCTGTGACAGCATGCCAAAGCAC	13220
QY	721	GCGTGTCTTGTGCTGTGTGTGACACTACATATAGGCGCTGATGCTGACTTG	780
DB	13221	AGCTTGTGCAAGCGCGGAGGAGCATGACAGATCTTTGGCCGAGCAGACCATC	13280
QY	781	GTCGCGCTCTTGCCGAAATGAGAGGGTATCTCTACGCAAACTGTATCTGAGTA	840
DB	13281	GCCACCCCGTTGGCGAATGAGAGAGGCTGTTTACGACACACTGTGATCTGCCGCT	13340
QY	841	CGCATCTTGTCTAAATGGCGGACAGCCCTGTGTCTATTATCCCGTCCGACATTACT	900
DB	13341	TTGATCTTGCCTCAAGTCGCGCGGACCCGCGAGGCGGCACTACTGCGCCTGTGACGTACA	13400
QY	901	CGCTTGTATAGATGCGCAGCCCTTAATTACCGGTAGTTGAAT	944
DB	13401	CGCCTGATGTTCAACCCCAATCCCAACCTTGTGCTGGAAT	13444

RESULT 7  
AY487512 999 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A9 nitrilase (BD5168) gene, complete  
DEFINITION cds.  
ACCESSION AY487512  
VERSION AY487512.1 GI:40890228  
KEYWORDS ENV,  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
1 (bases 1 to 999)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)

REFERENCE 2 (bases 1 to 999)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1. .999  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A9"  
/environmental\_sample  
1. .999  
/locus\_tag="BD5168"  
1. .999  
/locus\_tag="BD5168"  
/locus\_tag="BD5168"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97459.1"  
/db\_xref="GI:40890229"  
/translation="MKTTVTYACVQAAPVFMDEGTIDKTTILISEAAQKGAELIAFP  
ETWIPGYPMFLMNSPATMPLVQYHONSJLVDSQAKRIADARONNITVALGFS  
RDHGLYIAQWLIGSDGETIGIRKLKATHERTLFGESDSSLTWETPLGNVGLC  
CWEHLQPLSRYAMYSQHEBIHIAWPSLSLYTSATALGPDVNTAASRLYAAEGQCFV  
IAPCAVVSDEMIDLCPPDDRRALLSAGGHARITYGPDGRELVTPLGENEGLLIAEL  
DSAITPAKLAADPVGHYSRPDVTRLFPNSANKTVIKRHSPELIIAEQTAESEEE"

ORIGIN  
Query Match 39.4%; Score 399.6; DB 3; Length 999;  
Best Local Similarity 64.0%; Pred. No. 2.1e-107;  
Matches 603; Conservative 0; Mismatches 339; Indels 0; Gaps 0;

OY 1 ATGAAAGAAGCTATCAAGTGCCTGCGTGCAGCGCCCGCATCTACATGATTGGAG 60  
Db 1 ATGAAACAACGCTACCGCTTGCGTGCAGCGCCCGCATCTATTATGATTAGAA 60  
OY 61 GCGACGCTGACAAACCATGAGTTGATGAGAAGACAGCAGCTAATATGCTCTG 120  
Db 61 GGCACCGTAGATAAAACATCACCCCTCATCTCTGAAGCCGACAGAAAGCGCGAGCTC 120  
OY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACA 180  
Db 121 ATGCTTTTCCGAGACTTGATACCCGTTACCCGTTCTTATGCTGAAGTCCGCC 180  
OY 181 GCATGGGCAATGCAATTGTGACGCCAATACCATGAGAATCTCATTTGAGTTGAGGCCCT 240  
Db 181 GCCACAATATGCCCTGCTTTATCAGTATCATCAGAACTCTCTGCTGAGACGTAAC 240  
OY 241 CAAGCTAAGCGCATTTGATGACAGCCAGCGGTTGGGAATCATGTACCCCTGGGATG 300  
Db 241 CAGCGGAAGCGAATTCGATGCGGACAGCGCAGAAATACATCACTGCTGCTGCGCTTC 300  
OY 301 AGTGAACGGGTGCGTGCAACCTTTACATCAGTCAGTGGTTCATAGCGGATATGTTGAC 360  
Db 301 AGCGAAGCGATCATGGAAGCTCTATATCGCACAGTGGCTGATTGGCAGCGAGGGAG 360  
OY 361 ACCATTGGGGCCCGCGAAAGTGAACCTTCTTTGTTGAACGTACTTTGTTGGCGAA 420  
Db 361 ACCATTGGCATCCGCGCAAGCTCAAGGCCAGCAGTGGAGCGTACGCTGTTGGCGAA 420  
OY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGCTGTTGGAAGGCTGGTGGCTTATGC 480  
Db 421 AGCGACGGCTCCTCCGTGACCACTGGGAGACACCTCTGGGTAACTGGGGCCCTCTGC 480  
OY 481 TGTGGAGACACTTCAACCGCTAACAAATAGCTTTGTATGACACAAATGAGAGATT 540  
Db 481 TGCTGGAGACACTGACCGCGCTGTCCGCTATGCAATGTATCCAGCATGAGAGATC 540

OY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCTTAATGCGCGGAAAGCCCTGGGCCCT 600  
Db 541 CACATCGCTGCTGGCCAGTTTCAGTCTCTACACAGTAGCAAGCGCGCACTGGTCTCT 600  
OY 601 GATGTCAATGATAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
Db 601 GACGTCAATAGCGGCTTACGCTTATGCGCGAGGGGAGTGGCTTCTGATAGCC 660  
OY 661 TCGTGCCTGCTTTTCAACATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 CCGTGCCTGCTTTTCTGATGAATGATGATTTACTCTGCTGATGATGACCGGAGA 720  
OY 721 GCGTGTCTGCTGCTGCTGCTGACACATCACTATCATAGGCGCTGATGCTGACTTG 780  
Db 721 GCGTACTCAGTGCAGGAGGAGGACATGCCCTATTATTACGCGCCGAGCGAAGAACTC 780  
OY 781 GTCGCGCTTTCGCGAAATGAAGAGGTATTCTTACGCAAACTTGATCCTGGAGTA 840  
Db 781 GTCACCCCTCTCGGGGAAATGAGAGAGACTGCTTATCGTGAAGTCACTCTGCTGCG 840  
OY 841 CGCATCCTTGTAAATGCGCGCAGACCTGCTGTCATTATTCGGTCCGACATTACT 900  
Db 841 ATTACCTTGTCAAACTGCGCGCAGACCCGTTGGCCACTATTCGCTCTGACGTGACC 900  
OY 901 CGCTGCTAATGATCGCAGCCCTAATTAATCCGCTAGTTGAA 942  
Db 901 CGCCTCCTTTTAATCCTTACGCCCAACAGACTGTGATTAA 942

RESULT 8  
AY487492  
LOCUS 1002 bp DNA linear ENV 05-APR-2004  
DEFINITION Uncultured organism clone 2A10 nitrilase (BD7615) gene, complete  
cds.  
ACCESSION AY487492  
VERSION AY487492.1 GI:40890188  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
AUTHORS  
1 (bases 1 to 1002)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1002)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1. .1002  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A10"  
/environmental\_sample  
1. .1002  
/locus\_tag="BD7615"  
1. .1002  
/locus\_tag="BD7615"  
/locus\_tag="BD7615"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97439.1"  
/db\_xref="GI:40890189"  
/translation="MKTTVTYACVQAAPVFMDEGTIDKTTILISEAAQKGAELIAF"



ORIGIN

PETWIPGYPMFLWLNBPATNMLVYOHNSLVLDSAQAKRIADAQONNITVVLGFS  
ERDHGSLYISQWLIGSDGETIGIRKLKATHVERTLFGESDSSLTWTETPLGNVGL  
CCMEHLQPLSRYAMYSQHEIHIAAMPSFSLYTSATPALGPDVNTAABEGQCF  
VLAPCAVSDSEMIDLCPDDRRALLSAGGSHARIYGPDRRELVPTLGENEGILLIAE  
LDSAAITFAKLAADPVGHYSRPDVTRLFLFNPANKTVIKRHSPELIIAQAPEEEE"

Query Match 39.1%; Score 396.4; DB 3; Length 1002;  
Best Local Similarity 63.8%; Pred. No. 1.9e-106;  
Matches 601; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY	1	ATGAAAGAACTATCAAGTCCGCTGCGCAAGCCGCCGATCTTACATGATTTGAG	60
DB	4	ATGAAACAACCTGTTACCGTTGCTGCGTTACAGCCGCCCGTATTATGATTTAGAA	63
QY	61	GGACCGGTGACAAACCATGAGTTGATGAGAAGACGACGTAATATGCTCGCTG	120
DB	64	GGCACCATAGATAAAGCATCACCTCATCTCTGAAGCCGCACAGAAAGCCGAGCTC	123
QY	121	ATGCGCTTTCGGAAACTTGATTCAGGCTACCCATGTTTCTTGCTTGACTCACA	180
DB	124	ATCGCTTTCGGAGACCTGATACCCGGTTACCCGTTCTTATGCTGAACCTGCC	183
QY	181	GCATGGGCAATGCAATTTGTACGCCAATACCATGAACTCATTTGAGTTGATGCCCT	240
DB	184	GGCACAAAATATGCCCCCTGTTTATCAATATCATCAGAACCTCTGTGCTGACAGTCC	243
QY	241	CAAGCTAAGCGCATTTAGATGACAGCCAGCGTTGGAAATCATGTGACCCCTGGGATG	300
DB	244	CAGCGAAGCGAATGCGGATGTGACACAGAGAATAACATCATCTGCTTGGAATTC	303
QY	301	AGTGAACGGGTGCGTGACACCTTTACATCAGTCACTGTTTATAGCGGATAATGTGAC	360
DB	304	AGCGAGCGGATCATGGAAGCTCTATATCTCACAGTGGCTGATTTGGACGCGGGAA	363
QY	361	ACCATTTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAACTTCTTGTTCGGGAA	420
DB	364	ACTATTGGCATCCGGCGCAAGCTCAAGGCCACACACGTGAGCGGTGTTTCGGGAA	423
QY	421	GGGGATGTTTCATCGCTAGCGTTTTCGAGACGCTGTGTTGAAAGGCTGGGCTTATGC	480
DB	424	AGCGACGGCTCCTCCCTGACCACTGGAGACACCTTGCGTTAACGTGCGGCCCTCTGC	483
QY	481	TGTTGGAGACACCTTCAACCGCTAACAAATACGCTTGTATGACAAATAGAGATTT	540
DB	484	TGCTGGAGACCTGACGCGCTGTCTCGCTATGCGATGTATTTCCAGCATGAGAGATC	543
QY	541	CATTGTGCGGCTTGCCGAGCTTTAGCCTTTATCCTAATGCGCGGAAAGCCCTGGGCTT	600
DB	544	CATATCGCTGCGTGCCCGAGCTTCAAGTTTACACCAAGCGCAACTGCCGCTCGGACTT	603
QY	601	GATGTCAATGTAGCGGCGCTCTGCAATCTATGCCGTTGAAGGGCAATGCTTCTGTAAGCG	660
DB	604	GACGTCAACACGCGCGCTTCAAGCTCTATGCGCGGAGGGGCAAGTGTCTGTTAGCC	663
QY	661	TCGTGTGCGCTGTTTCAATTCATGATGATATGCTTTGTACAGATGACGAAAGCAT	720
DB	664	CCATGTGCGGCTGTTTCTGATGAGATGATTTACTCTGTCTGATGATGACCGGAGA	723
QY	721	GCGTTGCTTCTGCTGCTGCTGACACTCAGTATCATAGGGCTGATGGTGTGACTTG	780
DB	724	GCGTACTCAGTGCCGAGGGGGACATGCCGCTATTACGGAACCTGATGAGAGAACTC	783
QY	781	GTCGGCGCTCTTGCCGAAATGAAGGGTATTTCTTAAGCAACCTTGATCCTGAGTA	840
DB	784	GTCACCCCTCTCGGGGAAATGAGGAAGACTGCTTATCGCTGAGCTCGACTGCTGCG	843
QY	841	CGCATCCTTGTAAATGGGGGACAGACCTGCTGTGCTATTATCCCGTCCGAGATTACT	900
DB	844	ATCACTTTTGCCAAACTGGGGGACAGATCCCGTAGGCCACTATTCCCGCCCTGAGTGACC	903
QY	901	CGCTTGCTAATAGATCGAGCCCTAAATTACCGGTAGTTGAA	942

DB 904 CGCCTCTTTTAAATCCTTCAGCCACAAGACTGTTAATAA 945

RESULT 9  
AY487427  
LOCUS 1038 bp DNA linear ENV 05-APR-2004

DEFINITION Uncultured organism clone 2A12 nitrilase (BD7611) gene, complete cds.

ACCESSION AY487427  
VERSION AY487427.1 GI:40890058

KEYWORDS ENV.

SOURCE uncultured organism

ORGANISM uncultured organism

REFERENCE unclassified; environmental samples.

AUTHORS 1 (bases 1 to 1038)

Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.

Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)

2 (bases 1 to 1038)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.

Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Place, San Diego, CA 92121, USA

FEATURES  
source location/Qualifiers

1..1038 /organism="uncultured organism"

/mol\_type="genomic DNA"

/db\_xref="taxon:155900"

/clone="2A12"

/environmental\_sample

1..1038 /locus\_tag="BD7611"

1..1038 /locus\_tag="BD7611"

/EC\_number="3.5.5.7"

/codon\_start=1

/evidence=experimental

/product="nitrilase"

/protein\_id="AAR97374.1"

/db\_xref="GI:40890059"

/translation="MAIEHPYRYVAVOAPEFLNLEATVDKTLIEBAARGASLI  
AFPEWIPGYPMFLWLNBPATNMLVYOHNSLVLDSAQAKRIADAQONNITVVLG  
FSEKDGSLYIAQAILSPGKTIATRKLPKTHVERAIFEGDGSDAVHDTKLRVG  
ALCWEHIDPLSKYAMYAQNEQVHIAAMPSFSLYDAAYALGPEVNNASRLYAVEGQ  
CFVAPCATVSQKIMIDMLCETPEQQLKPGGGHAQIYGPDRSLADPLPDAEGLLY  
ADIDLAAITLAKAAADPAHYSRPDVTQLLLDRNPKPRVHAKPGSANNSSPGMRAY  
EHTELEBEGQA"

ORIGIN

Query Match 35.9%; Score 364.2; DB 3; Length 1038;  
Best Local Similarity 62.2%; Pred. No. 7.7e-97;  
Matches 573; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

QY	18	GGTGCGCTGCTGCAAGCCGCCGATCTACATGATTTTGAGGCGACGCTGACAAAC	77
DB	27	GCTGCTGCGCTTCAAGCGCGCGGACAGAGTTCTTAATCTGAGGCAACCGTTGACAGAC	86
QY	78	CATTGAGTTGATGAGAAGACGACGTAATAATGCTCGTGTGATCGCTTTCCGAAAC	137
DB	87	GATTGCGCTTATCAAGAGAGCGCCGCCGAGCGGCAAGTCTCATTTGCTTCCCTGAAC	146
QY	138	TTGATTTCCAGGCTAACCATGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGCAATT	197
DB	147	TTGATACCTTGATATCCATGTTTGGCTTGCTGCTGCGCCGATCTGGGCAATGAATTT	206
QY	198	TGTAGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCAATTTC	257



Db	207	CATCCAGCGTACCATGACAACTCGATGATGACGCGCCAGTTTGAAGCTATTGC	266
QY	258	AGATGAGCAAGCGGTTGGGAATCATGTACACCTGGGGATGATGAACGGGTGGTGG	317
Db	267	GCAAGCAGCATCGCGTTGCAACATCACTGTGTGCTTGGCTTAGCGAAGACGACGG	326
QY	318	CACCCCTTACATCACTGAGTGTGTTATAGCGCATATGCTGACACCATTTGGGCCCGCG	377
Db	327	AAGCCTGTATATGCCAGGCCATCCTGAGCCCTGAAGGAAGACCATCGCCACGGCTCG	386
QY	378	AAAGTTGAACCTACTTTTGTGTAACGCTACTTTGTTCCGCCGAAGGGGATGTTTCATCGCT	437
Db	387	CAAGCTGAACCCCACTCATGTGCAACGCGCGATCTTCGGCGAAGGCGACGGCAGCACCT	446
QY	438	AGCGGTTTTCGAGACGCTGTGTGGAAGGCTGGGTGCTTATGCTGTGGAGACCTTCA	497
Db	447	GCGAGTTCAACGACCAAGCTCGGCAAGGTGGCGCCCTTTGCTGTGGAGCATCTTCA	506
QY	498	ACCGCTAACAAATAAGCTTTGTATGACAAATGAAGAGATTCAATTGCGGCTTGCC	557
Db	507	GCCACTTTCAAATACGCGATGTATGCTCAGAACGACAGGTCATATCGCCGCTGGCC	566
QY	558	GAGCTTTAGCCCTTATTCCTAATGCGGCGAAAGCCCTGGGCTGATGTCAATGATAGCGGC	617
Db	567	CAGTTTTCCTTACGTGACGCGGCTTACGCGCTTGGGCCAGAGTGAACACGCCGC	626
QY	618	CTCTGAACTATGCGGTTGAAGGCAATGCTTCTACTAGCGTCTGTGCGCTGCTTTC	677
Db	627	GAGTCGGTGTATGCGGTGAGGGCCAGTGTGCTGCGGCTTGCGCAACGCTTTC	686
QY	678	ACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTGGCTGG	737
Db	687	GCAAAAGATGATCGATATGCTTTGCGAGACACCCGACACAAAGCGCTTGAAGCCGGG	746
QY	738	TGTTGACACTCAAGTATCATAGGGCTGATGTTGTACTTGTGCGGCTTGTCCGA	797
Db	747	GGGTGTCACGCGCAATCTACGGTCCCGACGACGATCGCTGGCCGATCCGCTGCC	806
QY	798	AAATGAAGAGGATTTCTTAACGCAACCTTGATCTGAGTACGCATCCTTGCTAAAT	857
Db	807	GGAACGCGAGGGGCTGTGTATGACAGACATTGACCTTGACGCCATCACCTCGCGAAAGC	866
QY	858	GGCGGACAGACCTGTGCTCATTTATCCCGTCCCGACATTACTCGCTTGCTAATAGATCG	917
Db	867	AGCTGACAGTCTGTGCGCATTAATCTCGCCCTGACGTGACACAACTGTGCTTGACCG	926
QY	918	CAGCCCTAAATTACCGGTAGT	938
Db	927	CAATCCAAAGCCCGTGTGT	947

RESULT 10  
AY487546 1011 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A16 nitrilase (BD5295) gene, complete cds.  
DEFINITION  
ACCESSION AY487546  
VERSION AY487546.1 GI:40890296  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
AUTHORS  
1 (bases 1 to 1011)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1011)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,

TITLE		Kretz,P.L., Burk,M.J. and Short,J.M.	
JOURNAL		Direct Submission	
SUBMITTED		(26-NOV-2003) Bioinformatics, Diversa Corporation, 4955	
DIRECTORS		Place, San Diego, CA 92121, USA	
FEATURES		Location/Qualifiers	
SOURCE		1. 1011	
		/organism="uncultured organism"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:155900"	
		/clone="2A16"	
		/environmental_sample	
gene		1. 1011	
		/locus_tag="BD5295"	
CDS		1. 1011	
		/locus_tag="BD5295"	
		/EC_number="3.5.5.7"	
		/codon_start=1	
		/evidence=experimental	
		/transl_except=(pos:1, .3,aa:Met)	
		/product="nitrilase"	
		/protein_id="AAR97493.1"	
		/db_xref="GI:40890297"	
		/translation="MGISHPKFAAVVQAPFLDLDGVERAVSLIGQAAEAGQOLI AFPETWIPGYPMHTWLGSPPAMEKEFVQRYFDNALRHGSPQERISGAAEHKIMVS LGFAERDGLYLIAQWLIIGPGQTIISRRKLKPTHERVTEGEGSDLSVHDTALGR IGSLCEWEHLQPLSKMYAQNQIHIGAMPSPSLYQFPANALSPEVNIAVSRVAYE GOCFLAPCATVSDAMIEITLCPDPEKQGLIRAGGHAALFGDGLPTPTVADTYEGL LYAELDLGVISIAXSAADPAGHSRPDVTRLLNQTPSKRYQNMVLPLETVTEPEGPV QP"	
ORIGIN			
Query Match		31.5%; Score 319; DB 3; Length 1011;	
Best Local Similarity		60.1%; Pred. No. 2.3e-83;	
Matches		551; Conservative 0; Mismatches 360; Indels 6; Gaps 1;	
QY	14	TCAAGTGCCTGCTGCAAGCCGCCGATCTACATGATTGGAGCGACGGTGACA	73
Db	23	TCAAGCTGCGTCTGTACAGCCGGGCTGCTTCTCGACCTGACGGCGCGTGAAC	82
QY	74	AAACCATGAGTTATGGAAGAAGCAGCAGTAATATGCTGCTGATGCGCTTCCGG	133
Db	83	GAGCGTGTGCTCATCGCCCAAGCAGCGCGAAGGGCAGCAGCTGATGCTTCTCG	142
QY	134	AACTTGATTCAGGCTACCCATGTTCTTGCTTACTCACCAGCATGGCAATGC	193
Db	143	AAACGTGATTCGCCGTATACCGTGCAACCTGCTTGCAAGCCCGGGTGGCAATGG	202
QY	194	AA-----TTGTACGCCAATACCATGAGACTCATTTGAGTTGATGGCCCTCAAGCTA	247
Db	203	AAAAAGGCTTTGTCCAACGATATTTGACACAGCGCTTGCGGCATGTTCTCCGCAAGCCG	262
QY	248	AGCGATTTCAGATGCAGCCAGCGGTTGGGAATCATGTGACCCCTGGGATGATGAAC	307
Db	263	AGCGAATCTCCGGGGCTGCGGGAGCACAAGATTATGTGTGCTTGGGTTGCCGAAC	322
QY	308	GGTGGTGACCCCTTACATCAGTCACTGTTTATAGCGGATATGATGACACCATTTG	367
Db	323	GCGATGAGGACGCTTTATATCGCGCAGTGGCTCATCGGACCCGACGGCCAAACTATCT	382
QY	368	GGGCGGCGGAAAGTTGAAACCTACTTTTGTGAACGTAATTTGTTCCGCGAAGGGATG	427
Db	383	CACGACGGCGAAGCTTAAGCCGACTCAGTCAGCGCACTGTATTTGGCGAGGAGACG	442
QY	428	GTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGAAGGCTGGGTGCTTATGCTGTTGG	487
Db	443	GAAAGCATCTCTCCGTGATGATACGGCGCTTGAAGTATCGCTCATCTTGTGCTGGG	502
QY	488	AGCACCCTCAACCGCTAACAAATACGTTGTATGCACAAATGAAGAGATTCAATTGTG	547
Db	503	AGCATTTGCAACCGTTGTGAATACGATGTACGCCAGATGAACAGATTACATTG	562
QY	548	CGGCTTGGCGAGCTTACCTTATCTTAATGCGCGGAAAGCCCTGGGCGCTGATGTCA	607

Db 563 GCGCATGGCCAGCTTTTCGTATACCAGCCATTGGCAATGCGCTGAGTCCCGAAGTCA 622  
QY 608 ATGTAGCGCCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCGTGTG 667  
Db 623 ATATCGCAGTAAGCCGCGTGTACGCCGTGAAGGCCAGTGTCTTCTCGCGCGTGC 682  
QY 668 CGCTCGTTTCAATCCATGCATGATATGCTTTGTACAGATGACGAAAAGCATGCGTTGC 727  
Db 683 CGACGGTTTCGACGCCCATGATCGAAACACTGTGCGATACGCCCGAAAAGCAGGACTGA 742  
QY 728 TTCTGGCTGTGTGTGACACTCAAGTATCATAGGGCTGATGTGTGACTTGGTCGCC 787  
Db 743 TTGGGGCGGTGGCGGACGCCGATCTTGGGCCAGATGGAAGTCTGCTGACGCCCTA 802  
QY 788 CTCTTGCCGAAAATGAAGGSTATTTCTCTACGCAAACTTGATCCTGAGTAGCATCC 847  
Db 803 CGGTAGCGGATCTTACGAGGGCTGCTGTATGCAGAATCGACCTCGCGCTCATTTTGA 862  
QY 848 TTGCTAAATGCGCGCAGACCCCTGTGTCTATTATCCCGTCCGACATTACTGCTTGC 907  
Db 863 TCGCAAGAGTGCAGCGGACCCCGCGCCACTATTCGGCGGCAGATGTCAACGCGCTTC 922  
QY 908 TAATAGATCGAGCCCT 924  
Db 923 TATTGAATCAGACGCCCT 939

RESULT 11  
AY487438 1026 bp DNA linear ENV 05-APR-2004  
LOCUS AY487438 1026 bp DNA linear ENV 05-APR-2004  
DEFINITION Uncultured organism clone 2A14 nitrilase (BD7266) gene, complete  
cds.  
ACCESSION AY487438  
VERSION AY487438.1 GI:40890080  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE unclassified; environmental samples.  
1 (bases 1 to 1026)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1026)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
JOURNAL Directors Place, San Diego, CA 92121, USA  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
FEATURES  
SOURCE  
1. 1026  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A14"  
/environmental\_sample  
1. 1026  
/locus\_tag="BD7266"  
1. 1026  
/locus\_tag="BD7266"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97385.1"  
/db\_xref="GI:40890081"  
/translation="MLSPVTQYRAAAVQAAPSLDLDRTEKTIATIEQAAEQDRLI  
AFPETWIPGYPLMIWLGPAMGRFVQRYFENSLVRGSKOMNAIDARHRMTVVVG  
FSERAGSLVMGOAIFGPEGELIARRKLKPTHAERTVFGEGDGHLLAVYETGVGRIG

ALCWEHIIQPLSKYAMYANEQVHVASWPCFSLYRGMAYALGDEVNTAASQVYAVEGG  
CYVLAASCLVVTPEILKVLIDTPDKPELLLAGGFSMIFGPDGRALAOPLPETEGLVT  
AEIDLGAIATALAKAAADPAGHYARPVDVTRLINPRPAARVEALGPRFEVQSEQAEPT  
QPAEAD"

ORIGIN  
Query Match 30.7%; Score 311.4; DB 3; Length 1026;  
Best Local Similarity 59.0%; Pred. No. 4.3e-81;  
Matches 534; Conservative 0; Mismatches 371; Indels 0; Gaps 0;  
QY 19 GTCGCTCGCTGCAAGCCGCCGATCTACATGATTTGAGGCGCAGCAAAACC 78  
Db 28 GCCCGCGGTGAGGCGCGGCCCATCTTTCTCGATCTGACCCGACCGTCGAGAAAGC 87  
QY 79 ATTGAGTTGATGAAGAAGACGACGCAATTAATGCTGCTGTGATCGCTTTCCGAAACT 138  
Db 88 ATCGGATCATGAGCAGCGCGCCGACAGATGTGCGCTGATGCGTTTCGGAACC 147  
QY 139 TGGATTCCAGGCTACCCATGTTTCTTGGCTTGACTCACGACGATGGCAATGCAATTT 198  
Db 148 TGGATTCCCGGCTATCCGCTCTGATCTGCTCGGCTCGCCGCTGCGGATGCGCTTC 207  
QY 199 GTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGATTTC 258  
Db 208 GTGACGCGCTATTTCGAGAACTCGCTGTGCGCGGCAGCAAAACAGTGGAACGCGATCGCC 267  
QY 259 GATGACGCCAAGCGGTTGGGAATCATGTGTCACCTGGGATGATGAACGGTCCGTGGC 318  
Db 268 GATGCGGCGCGGCCACCGCATGACCGTCTGCTGCTTACGAGCGCGGGAAGGC 327  
QY 319 ACCCTTTACATCAGTCAAGTGTTCATAGGCGATATGTGTACACCATGGGCGCCGCA 378  
Db 328 AGCCTTACATGCGGCCAGCGCATCTTGGCCCCGAAGCGAGCTCATCGCGCGCGCCG 387  
QY 379 AAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGGATGTTTCATCGCTA 438  
Db 388 AAGCTCAAGCGCACACAGCCGAGCGAAGGTTGTCGGAGGCGACGCCACTTG 447  
QY 439 GCGGTTTTCGACAGCTGTGTGAAGGCTGGGTGCTTATGCTGTTGGAGCACCTTCAA 498  
Db 448 GCCGTTTACGAGACGGGCGTGTGCGCATCGGCCCTCTGCTGTGGAGCACATCCAG 507  
QY 499 CCGCTAACAAATAACGCTTTGTATGCACAAATGAAGAGATTCAATGTGCGGCTTGGCCG 558  
Db 508 CCGCTCTGAAATACGCGATGTATGCGGCCAACGAACAGGTGATGTGCTCTGTGCGCG 567  
QY 559 AGCTTAGCCTTATCTTAATGCGGCAAGCCCTGGGCGCTGATGCAATGTAGCGGCC 618  
Db 568 TGCTTACGCTTATTCGCGCATGCGCTATGCGCTCGGCGGAGGTGAACACGCCGCG 627  
QY 619 TCTGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGCGTGTGCGCTGTTCA 678  
Db 628 AGCCAGGTCTACGCGGTCGAGGCGGCTGCTACGTGTGCTCTCTGCTGCTGACA 687  
QY 679 CAATCCATGATGATATGCTTTGTACAGATGAGAAAGCATGCGTGTCTTGCGTGT 738  
Db 688 CCCGAGATCTGAAAGGTGCTGATCGACACGCCCGACAAAGAGCGGTGCTGCTGCGCG 747  
QY 739 GGTGACACTCAAGTATCATAGGGCTGTGTGTGACTTGTGCGGCTCTTGCCGAA 798  
Db 748 GGGGGTTCTGATGATCTTTCGCGCCGACGCGCGCGCTCGCCACGCGCTGCGGAG 807  
QY 799 AATGAAGAGGTATTTCTTACGCAAACTTGATCTGAGTACGATCTTGTCTAAATG 858  
Db 808 ACCGAAGAGGCTCTGTCACGCGCGAGATCTCGCGCGATGCGCTCGCAAGGCC 867  
QY 859 GCGCAGACCTGTGTCATTAATCCGTCGCCGACATTAATGCTGCTGCTAATAGATGCG 918  
Db 868 GCGGCGGATCCCGCGCCATTAACGCGGCGCCGACGTGACGCGGTTGTGTAACCCG 927  
QY 919 AGCCC 923  
Db 928 CGCCC 932

RESULT	12
AY487522	1077 bp DNA linear ENV 05-APR-2004
LOCUS	AY487522
DEFINITION	Uncultured organism clone 2A17 nitritease (BD5263) gene, complete cds.
ACCESSION	AY487522
VERSION	AY487522.1
KEYWORDS	GI:40890248
SOURCE	ENV.
ORGANISM	uncultured organism
REFERENCE	uncultured organism
AUTHORS	unclassified; environmental samples. 1 (bases 1 to 1077) Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M. Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)
TITLE	2 (bases 1 to 1077) Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.
JOURNAL	Direct Submission
REFERENCE	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA
AUTHORS	Location/Qualifiers 1. .1077 /organism="uncultured organism" /mol_type="genomic DNA" /db_xref="taxon:155900" /clone="2A17" /environmental_sample 1. .1077 /locus_tag="BD5263" 1. .1077 /locus_tag="BD5263" /EC_number="3.5.5.7" /codon_start=1 /evidence=experimental /product="nitrilase" /protein_id="AAK97469.1" /db_xref="GI:40890249"
FEATURES	/translation="MALTHPKLVAAVQAAPFLDVDADKAVRLIDEAANGSSLV APPETWIPGYPFWMUGSPAWAIMRGFVSRYPDNSLSYDSROAERLRDAKRHLTV MGLSERAGSLYTAOWIIGPNETVAORRKPTHAERTVFEGDGSHTAHNLPIGR LGALCSWEHLOPLSKYAMYAQNEIHVAAMPESLYDPFAHALGAEVNNAASQIYAVE GSCFVAPCAVISDEMIDMLCDTPDKHOLIHYGGFTVIYGPDGARIGDKLPDQEGI VYADIDLGMIPIAKAADPAIGHYARPDTVTRLLENNRPANNRVETTLVPVDQVIDARV EAAAPGARPATGNEDPPAKMAAE"
gene	
CDS	
ORIGIN	
Query Match	30.5%; Score 309.4; DB 3; Length 1077;
Best Local Similarity	59.8%; Pred. No. 1.7e-80;
Matches	539; Conservative 0; Mismatches 356; Indels 6; Gaps 1;
Db	14 TCAAGTTCGCTGCCTGCGTGAAGCCGCCCGATCTACATGATTGGAGCGCAGCGTGACA 73
QY	11
Db	23 TGAAGTCGCCCGCGTGCAGGACGTCCCGCTTCTCGATGTCGATGCCGACGTGACA 82
QY	74 AAACCATGAGTTGATGGAAGAAGCAGCACGTAATAATGCTCGTCTGATCGCCTTCCGG 133
Db	83 AAGCGTGCGGCTAATCGACGAAGCGGCAGCAAACGGCTCCAGTCTGGTGCAATCCCCG 142
QY	134 AAACCTGATTCAGGCTACCCTATGCTTTCTTTGGCTTGACTCACCAGCATGGGCAAT-- 191
Db	143 AGACCTGATCCCCCGCTATCCGTTTTGGATCTGGCTTGCTCGCCGGCCTGGGCAATCA 202
QY	192 ----GCAATTGTACGCCCAATACATGAGAATCTATTGGAGTTGGATGGCCCTCAAGCTA 247
Db	203 TGCGCGGGTTGTGTCTCGCTATTTCGATTAATTGCTCAGCTATGACAGCCGGCAGGCA 262

QY	248	AGCGCATTTTCAGATGCAGCCCAAGCGGTTGGGAATCATGATGTCACCCCTGGGGATGAGTGAAC	307
	263	AGGCGCTGCGGACGCGCGGAGCGCCACAAACTGACCGTGTCTATGGGCGCTGTCCGAGC	322
QY	308	GGGTGCGTGACACCCCTTACATCAGTCAGTGGTTTCATAGCGGATTAATGTTGACACCATTG	367
Db	323	GCGCGGCGGTAGCCCTTTACATCGCGCAGTGATCATTTGTTCCCAATGGCGAGACGTCG	382
QY	368	GGGCGCGGCGAAAGTTGAAACCTACTTTTGTGTAACGTACTTTGTTCCGCCAAGGGGATG	427
Db	383	CACAGCGGCGCAAGCTCAAGCCCAACCATGCGGAGCGCAACCGTCTTCGGCGAGGGTGAAG	442
QY	428	GTTTCATCGCTAGCGGTTTTCGAGACGTCTGTTGAAAGCGTGGTGGCTTATGCTTTGGG	487
Db	443	GCAGCCACTGCGCGGTACACCAATCTTCCATCGAGCGGCTCGGTGCGCTGTGCTGGG	502
QY	488	AGCACCTTCAACCGCTACACAAATACGCTTTGTATGCACAAATGAAGAGATTCAATTGTG	547
Db	503	AGCACCTTCAGACCGCTCTCCAAATACCGCATGTATCGCCCGAAGAGAGATCCACGTCG	562
QY	548	CGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGCGCAAGCCCTGGGCGCTGATGTCA	607
Db	563	CGGCATGGCGGCTCTTCTCGCTTACGACCCGTTGGCGACGCGCTCGGCGCGCAAGTCA	622
QY	608	ATGTAGCGGCTCTCGAATCTATGCGGTGAAGGGGCAATGCTTCTGTAAGCGTCTGTG	667
Db	623	ACAACGCAAGCAGCCAGATCTACGCGGTGAAGGTTCTGCTTTGCTGCGCGCATGTG	682
QY	668	CGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAAGCATGCGTTGC	727
Db	683	CGGTGATCTGCAGAAATGATCGATCTTATGTGCGATACCCCGACAAAGCATCAGCTTA	742
QY	728	TTCTGCTGTGTGTGACACTCAAGTATCATAGGCGCTGATGTGTGACTTGTGCGGC	787
Db	743	TTACAGTGGTGTGCGGCTTCAACGTGATCTATGCGCCGAGCGGTGCGCATGCGCGACA	802
QY	788	CTCTTGCCGAAATGAAGAGGGTATCTCTACGCAAACTTGATCCTGAGTAGCATCC	847
Db	803	AGCTCGCGCCAGATCAGGAAGCATGTCTATGCCGACATGATCTCGCATGATCCGGA	862
QY	848	TTGCTAAATGCGCGGACAGACCCCTGTGTCATTATTTCCCGTCCGACATTACTCGCTTC	907
Db	863	TGCGAAAGCTGCGCGCATCTTGTGCGGCCACTATGCGCGACCCGACGTTAACCCGCTTC	922
QY	908	T	
Db	923	T	
RESULT 13	AY487430	1047 bp	DNA linear ENV 05-APR-2004
LOCUS	AY487430		
DEFINITION	Uncultured organism clone 2A15 nitrilase (BD5338) gene, complete		
ACCESSION	AY487430		
VERSION	AY487430.1		GI:40890064
KEYWORDS	ENV.		
SOURCE	uncultured organism		
ORGANISM	unclassified; environmental samples.		
REFERENCE	1 (bases 1 to 1047)		
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		
TITLE	Exploring nitrilase sequence space for enantioselective catalysis		
JOURNAL	Appl. Environ. Microbiol.		
REFERENCE	2 (bases 1 to 1047)		
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		



TITLE	Direct Submission
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA
FEATURES	location/Qualifiers
source	1.1047
	/organism="uncultured organism"
	/mol_type="genomic DNA"
	/db_xref="taxon:155900"
	/clone="2A15"
	/environmental_sample
gene	1.1047
	/locus_tag="BD5338"
CDS	1.1047
	/locus_tag="BD5338"
	/EC_number="3.5.5.7"
	/codon_start=1
	/evidence=experimental
	/product="nitrilase"
	/protein_id="AAR97377.1"
	/db_xref="GI:40890065"
	/translation="MPTSKQFRVAAVQAAVFLDLLEGALISKGISLIEEASNGAKLIAFPETWIPGYPMWIMLDSFPAMGRFVORYFDNSIMLGSEQAKRMNQAANNKIYVMGYSERSGSLYMGOSIINDKETIFTRKIKPTHERTVFEGDGSILCYMDTEIGRVGAMCCWEHLQPLSKYAMYSDQEIHIASWPSFSLYRGAAYALGPDLNNAASQMYAAEGQC FVLAPCATVSKEMIEMLIDDPKPEPLLEGGFTMIYGPDRPLAKPLPENEGLLIYA DIDIAMI SMAKRAADPA GHYARPDVTRLFNSAPANRVEYINPASGPTESLKDMGKMQ MEAEQOKAALREMI"
ORIGIN	
Query Match	30.4%; Score 308.4; DB 3; Length 1047;
Best Local Similarity	58.9%; Pred. No. 3.4e-80;
Matches	531; Conservative 0; Mismatches 371; Indels 0; Gaps 0;
QY	14 TCAAGTGCCTGCGTCGAAGCCGCCCGCATCTACATGATTTGAGGCGACCGGTGACA
Db	20 TTAGAGTCGTCAGTTCAGCCGCCCGGATTTCTTGACCTGGAGGCGCAATTAACA
QY	74 AAACCATGAGTTGATGGAAGAAGCAGCAGTAATATGCTCGTGTGATCGCTTTCCGG
Db	80 AAGGCATCTCCCTCATTTGAGAGGCGCGCTTCCAATGAGCCAACTCATTTGCCCGG
QY	134 AAATTGATTCAGGCTAAGCCATGTTTCTTGCTTGACTCACAGCATGGGCAATGC
Db	140 AAAGTGGATTCGCGGTACCCCTGGTGATCTGGCTGACTCACCGCTTGGGCATGC
QY	194 AATTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCA
Db	200 GCTTTGTCAGCGCTATTTTGACAACTCGCTCATGTGGTAGTGAGCAAGCCAAAGCGCA
QY	254 TTTCAGATGAGCCCAAGCGGTGGGAATCATGTGCACCTGGGGATGAGTGAACGGGTGC
Db	260 TGAACCAAGCTGCCGCCAATTAACAAGATTACGTGTTGATGGGTTATAGCGCAACGAGTG
QY	314 GTGGCAACCTTTACATCAGTCAGTGGTTCATAGGCGATTAATGCTGACACCATTTGGGCC
Db	320 GCGGACGCTCTACATGGGCCAATCATTTATCAACGACAAGGTTGAAAAGATTTTAACC
QY	374 GCGGAAAGTTGAAACCTACTTTTGTGTAACGTACTTTTTCGGCGAAGGGGATGTTCAAT
Db	380 GCCGCAAACTCAAGCCAACTCATGTGAGCGTACCGGTTTGGGAGGAGACGCGACCC
QY	434 CGCTAGCGGTTTTCGAGACGTCGTGTTGAAGGCTGGGTGCTTATGCTGTTGGAGCAC
Db	440 ATCTTTGCGTATGATACCGAGATTTGGCGCGCTCGCGCGCATGTGCTGTTGGAAACATT
QY	494 TTCAACCGCTACAAAATACGCTTTGTATGCACAAAATGAAGATTCAATGTCGGCTT
Db	500 TGCAGCCGCTCAGCAAAATATGCAATGTATTCTCAGGATGAACAAATTCAATTGCCCTCT
QY	554 GCGCGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAG
Db	560 GCGCGAGCTTTTCGTTATATCGGGGGCGACCTATGCACTCGGCCCGGAACTGAACAAG

QY	614 CGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTGCCTACTAGCGCTGTCGCTCG
Db	620 CCGCAGCCAAATGTATGACGCCGAAGGCCAGTGTCTTGTCTTGCCCTTGGCCACCG
QY	674 TTTCACATCCATGATCGATATGCTTTGTAAGATGACGAAAAGCATGCTGCTCTCG
Db	680 TCTCAAGAGATGATCGAAATGCTGATAGATCCAGAAAGAGCCGCTTGTGCTGG
QY	734 CTGCTGTGACACTCACGTATCATAGGGCTGATGTTGTTGACTTGTGCGGCTCTTG
Db	740 AAGTGGCGGTTCAACCATGATTACGGCCCGATGGCGACCTTGCTAAACCGTTGC
QY	794 CCGAAATGAAGGGTATTCTCTACGCAAACTTGATCCTGGAGTACGCATCCTTGCTA
Db	800 CTGAAAACGAGGAGGGCTGCTATATGCCGATATTGACCTGGGCATGATTCAATGGCCA
QY	854 AATGCGGCGACGCTGCTGTGTCATTATTCGCTCCGACATTACTCGCTTGCTAATAG
Db	860 AGGCTGCCCGCGACCGCGCAGGTCACTACGACGCCCGGATGTCACTCGCTACTATTCA
QY	914 AT 915
Db	920 AT 921
RESULT 14	
AY487543	
LOCUS	1062 bp DNA linear ENV 05-APR-2004
DEFINITION	Uncultured organism clone 2A20 nitrilase (BD7753) gene, complete
ACCESSION	AY487543
VERSION	AY487543.1 GI:40890290
KEYWORDS	ENV.
SOURCE	uncultured organism
ORGANISM	uncultured organism
REFERENCE	1 (bases 1 to 1062)
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.
TITLE	Exploring nitrilase sequence space for enantioselective catalysis
JOURNAL	Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)
REFERENCE	2 (bases 1 to 1062)
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.
TITLE	Direct Submission
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA
FEATURES	location/Qualifiers
source	1.1062
	/organism="uncultured organism"
	/mol_type="genomic DNA"
	/db_xref="taxon:155900"
	/clone="2A20"
	/environmental_sample
gene	1.1062
	/locus_tag="BD7753"
CDS	1.1062
	/locus_tag="BD7753"
	/EC_number="3.5.5.7"
	/codon_start=1
	/evidence=experimental
	/product="nitrilase"
	/protein_id="AAR97490.1"
	/db_xref="GI:40890291"
	/translation="MGVAHPKYKVAAVQAAPAFLLDLASVEKAVRFIDEAGAAGARLI AFPETWIPGYPMWIMLGAPAWAIMGFVSRYPFNSLSYDSPOAEKLRRAAKRMNVV LGLSERDGSLYIAQWIIIGPDERTIAKRKLPTHAERTVFEEDGSHLAVHELDVGR LGALCCWEHLQPLSKYAMYQNEQVHVAAMPFSLYDPFAHALGAEVNNAAASKIYAVE GSCFVIAPCATVSKAMIDELCDTPEKHQFLHAGGFAVIYGPDAPIAAPLPDKEGL



LYADIDLGMISVAKAAADPAGHYARDVTRLLENNRPGYRVETMALPIDAETKAEAPA  
KPEPKAPNVAFPAFVQAAE"

ORIGIN

Query Match 30.3%; Score 307; DB 3; Length 1062;  
Best Local Similarity 59.0%; Pred. No. 8.9e-80;  
Matches 548; Conservative 0; Mismatches 375; Indels 6; Gaps 1;

QY	15	CAAGTGCCTGCGTGAAGCCGCCCGATCTACATGATTGGAGCGGCGTGACAA	74
DB	24	CAAAAGTGGCCCGTGAAGGACGCGCCGCTTTCTCGACCTGACGCGCTCGTGA	83
QY	75	AACCATGAGTTGATGAGAAGACGACACGTAATAATGCTGCTGATCGCCTTCCGGA	134
DB	84	GGCCGTCGTTTCATCGAAGCCGCGCGCCGCGCGCCGCTCATCGCCTTCCGGA	143
QY	135	AACTTGATTCCAGGCTAACCCATGTTCTTTGGCTTGACTCACACGATGGGCAAT--	191
DB	144	GACCTGATACCCGGTACCCCTGTGATCTGGCTAGCGCGCCGCTGGGCTATCAT	203
QY	192	--GCAATTTGTACGCCCAATACCATGAGACTCATGGAGTTGGATGGCCCTCAAGCTAA	248
DB	204	GCGCGGCTTCGTCCTCGCGCTATTTCGACAACTCGCTCAGCTACGACAGCCCGAGGCCGA	263
QY	249	GCGCATTTCAAGTCAAGCCAAAGCGGTTGGGAATCATGTCAACCTGGGGATGATGAACG	308
DB	264	GAACTCCGCGCCCGCCCAAGCGCAACAAGATGTGTGTGCTCGGCGCTTCGGAAGCG	323
QY	309	GCTCGGTGGCACCCTTTACATCAGTCAAGTGTTCATAGGCGATAATGTTGACACCATTG	368
DB	324	CGACGGCGGACGCTTTACATCGCGCAATGATTCGGCCCGGACGGCGAACCATCGC	383
QY	369	GGCCCGCGGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGGATGG	428
DB	384	CAAGCGCCGCAAGCTCAAGCCGACCCACGCGGAGCGGACCGGTTCGCGAAGCGACGG	443
QY	429	TTCATCGTAGCGGTTTTCGAGACGTCTGTTGAAAGCTGGGTGGCTTATGCTTTGGGA	488
DB	444	CTCGCATCTTGCGGTGACGAGCTTGATGTTGGCGGCTCGGCGGCTGTGCTGTGGGA	503
QY	489	GCACTTCAACCGCTAACAAATACGCTTGTATGACAAATGAGAGATTCTTGTGC	548
DB	504	ACACCTGCAGCCGCTGTCCAAATACGCCATGTATGCCAGAACAGACAGTGCATGTCCG	563
QY	549	GGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGGCGGAAAGCCCTGGGGCTGATGTCAA	608
DB	564	GGCTTGGCCGAGCTTTTCGCTTTAAGATCCGTTCCGCCACGCGCTCGGCGCGGAAGTGA	623
QY	609	TGTAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGCGCTCGTGTGC	668
DB	624	CAATGCGCGGAGCAAAATCTATGCGGTGAGGGCTCGTGTTCGTCATCGCGCGCTGCGC	683
QY	669	GCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCATGCGTGTCT	728
DB	684	GACCGTTTTCGAGCGATGATGACGAACCTCTCGATACGCCGAGACATCACTTCTCT	743
QY	729	TCTGCTGTGTGTGACACTGATCATAGAGGCTGATGTGTGATGACTTGTGTGCGCC	788
DB	744	GCATGCGCGGCGCGGCTTTGCGGTGATTTACGCGCCGACGCGCGCCGCTCGCGGCGCC	803
QY	789	TCTTGCAGAAATGAAGAGGTATTCTCTACGCAAACTTGATCCTGAGTAGCATCTCT	848
DB	804	GCTGCCCGCCGACAAAGGAGGCTTCTCTACGCCGACATGATCTCGGATGATTTCGT	863
QY	849	TGCTAAATAGCGGAGACCTGTGTGCTATTATCCCGTCCGACATTAATCTCGCTGTCT	908
DB	864	TGCCAAAGCGGAGCGGATCCGGCGGGCATTAATGACAGCCCGGACGTCAACCCGCTTCT	923
QY	909	AATAGATCGAGCCCTAAATTAACCGGTAG	937
DB	924	GTTCAACAATCGGCTGGGTATCGGGTCTG	952

RESULT 15  
AY487497  
LOCUS  
DEFINITION  
cds.  
uncultured organism clone 2A13 nitrilase (BD5086) gene, complete

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

AY487497 1041 bp DNA linear ENV 05-APR-2004  
AY487497  
AY487497.1 GI:40890198  
ENV.  
uncultured organism  
uncultured organism  
unclassified; environmental samples.  
1 (bases 1 to 1041)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1041)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
Location/Qualifiers  
1..1041  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A13"  
/environmental\_sample  
1..1041  
/locus\_tag="BD5086"  
1..1041  
/locus\_tag="BD5086"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97444.1"  
/db\_xref="GI:40890199"  
translation="MSEPMYKYRGAIVQAPVFLDLDRIVEKAIGLEQAKQDVRLI  
AFPTWIPGYFWIMILGAPAWMRVQRYFENSLVRGSKQOALDAARHGMYVAG  
YSEKAGSLYMGQALFGPDGLIARRKLPKTHAERTVGEKSHLAVHDTAIGRLG  
ALCMEHIQPLSKYAMYADEQYHVAWSPSFLYRGMAVALGPEVNTAASQIYAVEGG  
CYVLASCATVSPEMIKVLVDTPDKEMFLKAGGFAMIFGDGRALAEPLPETEGLLV  
ADIDLGMIALAKAAADPAGHYBRPDVTRLLENNRPGYRVETMALPIDAEPQNEKGDAPA  
LRVAVESAAAQ"

gene  
CDS

1..1041  
/locus\_tag="BD5086"  
1..1041  
/locus\_tag="BD5086"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97444.1"  
/db\_xref="GI:40890199"  
translation="MSEPMYKYRGAIVQAPVFLDLDRIVEKAIGLEQAKQDVRLI  
AFPTWIPGYFWIMILGAPAWMRVQRYFENSLVRGSKQOALDAARHGMYVAG  
YSEKAGSLYMGQALFGPDGLIARRKLPKTHAERTVGEKSHLAVHDTAIGRLG  
ALCMEHIQPLSKYAMYADEQYHVAWSPSFLYRGMAVALGPEVNTAASQIYAVEGG  
CYVLASCATVSPEMIKVLVDTPDKEMFLKAGGFAMIFGDGRALAEPLPETEGLLV  
ADIDLGMIALAKAAADPAGHYBRPDVTRLLENNRPGYRVETMALPIDAEPQNEKGDAPA  
LRVAVESAAAQ"

ORIGIN

Query Match 30.1%; Score 305.4; DB 3; Length 1041;  
Best Local Similarity 58.0%; Pred. No. 2.7e-79;  
Matches 540; Conservative 0; Mismatches 391; Indels 0; Gaps 0;

QY	19	GTCGCTGCGTGAAGCCGCCGATCTACATGGATTTTGAGGCGGAGTGAACAAACC	78
DB	28	GGCGCGGCGGTGACAGCCGCGCGGTGTTCTCGATCTGACCGCACAGTCGAGAAAGCG	87
QY	79	ATTGAGTTGATGAGAAGACGACACGTAAATGCTCGTCTGATCGCCTTTCGGAAACT	138
DB	88	ATCGGCTGATCGACGAGCGGCCAACGACGATGCGCTGATCGCATTCACAGAGACT	147
QY	139	TGATTCCAGGCTAACCATGCTTCTTTGGCTTGACTCACACGATGGGCAATGCAATTT	198
DB	148	TGATTCCCGGCTATCCCTTTTGATATGGCTGGCGCGCGCTTGGGCGATCGCTTC	207
QY	199	GTAAGCAATACCATGAGAATCTCATTTGAGTTGATGGCCCTCAAGCTAACGCCATTCA	258
DB	208	GTCACGCGCTATTTCAGAAATTCGCTCGTCGCGGCGACGAAGCACTGGCAGGCCCTGGCG	267
QY	259	GATGACGCAAGCGGTGGGAATCATGTCAACCTGGGATGAGTGAACGGGTGCTGGTGC	318



**THIS PAGE BLANK (USPTO)**





PS Claim 28; Page 85-86; 87pp; English.  
XX The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolysing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a DNA encoding nitrilase #2  
XX  
SQ Sequence 1014 BP; 237 A; 232 C; 280 G; 265 T; 0 U; 0 Other;

Query Match 100.0%; Score 1014; DB 4; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1014; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAGAAGCTATCAAGGTCGCTGCGTGCAGACCCCGCATCTACATGATTGGAG 60  
DB 1 ATGAAAGAAGCTATCAAGGTCGCTGCGTGCAGACCCCGCATCTACATGATTGGAG 60  
QY 61 GCGACGGTGACAAAACCATTTGATGTGATGGAAGACAGACAGTAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAAACCATTTGATGTGATGGAAGACAGACAGTAATATGCTGCTG 120  
QY 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCACCA 180  
DB 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTAGCCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
DB 181 GCATGGGCAATGCAATTTGTAGCCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAGATGCAGCCAAAGCGTTGGAAATCATGTCAACCCGTGGGATG 300  
DB 241 CAAGCTAAGCGCATTTCAGATGCAGCCAAAGCGTTGGAAATCATGTCAACCCGTGGGATG 300  
QY 301 AGTGAACGGGTCGGTGACACCTTTATCATCAGTCAGTGGTTCATAGCGATAATGGTGAC 360  
DB 301 AGTGAACGGGTCGGTGACACCTTTATCATCAGTCAGTGGTTCATAGCGATAATGGTGAC 360  
QY 361 ACCATTGGGGCCCCGGGAGTAAGTGAACCTACTTTTGTGAACGTACTTTGTTGGGAA 420  
DB 361 ACCATTGGGGCCCCGGGAGTAAGTGAACCTACTTTTGTGAACGTACTTTGTTGGGAA 420  
QY 421 GGGGATGCTTCATCGCTAGCCGTTTTCGAGACGTCGTGTGAAGCGTGGTGCTTATGC 480  
DB 421 GGGGATGCTTCATCGCTAGCCGTTTTCGAGACGTCGTGTGAAGCGTGGTGCTTATGC 480  
QY 481 TGTGGGAGCACCCTTCAACCGCTAACAAAATACGCTTGTATGCACAAAATGAAGAGATT 540  
DB 481 TGTGGGAGCACCCTTCAACCGCTAACAAAATACGCTTGTATGCACAAAATGAAGAGATT 540  
QY 541 CATTTGCGGCTTGCCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCTT 600  
DB 541 CATTTGCGGCTTGCCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCTT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
DB 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCAT 720  
DB 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCAT 720  
QY 721 GCGTGTCTTCTGGCTGTGTGACACACTACGATACAGGCGCTGATGGTGACTTGG 780  
DB 721 GCGTGTCTTCTGGCTGTGTGACACACTACGATACAGGCGCTGATGGTGACTTGG 780  
QY 781 GTCGCGCCTTTGCCGAAATGAAGAGGGTATTTCTTACGAAACCTTGATCCTGAGTA 840  
DB 781 GTCGCGCCTTTGCCGAAATGAAGAGGGTATTTCTTACGAAACCTTGATCCTGAGTA 840  
QY 841 CGCATCCTTGTAAATGCGGCGACAGCCCTGCTGTCATTATCCCGTCCCGACATTACT 900

DB 841 CGCATCCTTGTAAATGCGGCGACAGCCCTGCTGTCATTATCCCGTCCCGACATTACT 900  
QY 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCCGTAAGTGAATGAAGTGATCTTGGT 960  
DB 901 CGCTTGCTAATAGATCGACGCCCTAAATTACCCGTAAGTGAATGAAGTGATCTTGGT 960  
QY 961 CCTTACGCTTTGGGTAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGGA 1014  
DB 961 CCTTACGCTTTGGGTAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGGA 1014

RESULT 2  
ADCC24118  
ID ADCC24118 standard; DNA; 1014 BP.  
XX  
AC ADCC24118;  
DT 18-DEC-2003 (first entry)  
XX

DE DNA sequence (SeqID 385) encoding a nitrilase enzyme.

KW Gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KM enantiomer; chiral medicine.  
XX

OS Unidentified.  
XX

PN WO2003000840-A2.  
XX

XX 03-JAN-2003.  
PD  
XX

PF 15-MAY-2002; 2002WO-US015983.  
XX

PR 21-JUN-2001; 2001US-0300189P.  
XX

PR 30-JUL-2001; 2001US-0309006P.  
XX

PR 22-JAN-2002; 2002US-0351336P.  
XX

PA (DIVE-) DIVERSA CORP.  
XX

PA (MADD/) MADDEN D.  
XX

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
XX

PI Short JM, Burk M;  
XX

DR WPI; 2003-201417/19.  
XX

DR P-PSDB; ADCC24119.  
XX

PS Claim 1; SEQ ID NO 385; 560pp; English.  
XX

XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.  
XX

SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;

Query Match 99.8%; Score 1012.4; DB 10; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGCAGACCGCCCCGATCTACATGGATTGGAG 60  
Db 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGCAGACCGCCCCGATCTACATGGATTGGAG 60  
QY 61 GCGACGGTGCACAAAACCATTTAGTTGATGGAAGAAGCAGCAGCTAATAATGCTGCTG 120  
Db 61 GCGACGGTGCACAAAACCATTTAGTTGATGGAAGAAGCAGCAGCTAATAATGCTGCTG 120  
QY 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGTTCTTGGCTTGACTCACCA 180  
Db 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGTTCTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAATCATTTGGAGTTGGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAATCATTTGGAGTTGGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAAGATGCAGCCAAAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
Db 241 CAAGCTAAGCGCATTTCAAGATGCAGCCAAAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
QY 301 AGTGAACGGGTCCGTGGCACCCTTTACATCAGTCAGTGTTCATAGGCGATATGATGAC 360  
Db 301 AGTGAACGGGTCCGTGGCACCCTTTACATCAGTCAGTGTTCATAGGCGATATGATGAC 360  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAAGCTACTTTGTGGCGAA 420  
Db 361 ACCATTGGGGCCCGCGAAAGTTGAAACCTACTTTTGTGAAGCTACTTTGTGGCGAA 420  
QY 421 GGGGATGTTTCATCGTAGCGGTTTTCGAGACGTCGTGTGGAAGCGTGGTGCTTATGC 480  
Db 421 GGGGATGTTTCATCGTAGCGGTTTTCGAGACGTCGTGTGGAAGCGTGGTGCTTATGC 480  
QY 481 TGTGGGAGACACCTCAACCCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
Db 481 TGTGGGAGACACCTCAACCCGCTAACAAATACGCTTGTATGCACAAATGAAGAGATT 540  
QY 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGCT 600  
Db 541 CATGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGCT 600  
QY 601 GATGTCAATGTAGCGGCTCTGCAATCTATGCCGTTGAAGGCAATGCTTCTGACTAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTGCAATCTATGCCGTTGAAGGCAATGCTTCTGACTAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTGTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
Db 721 GCGTGTCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
QY 781 GTGCGGCTCTTGGCGGAAATGAAGAGGATTTCTTACGCAACCTTGATCTGAGATA 840  
Db 781 GTGCGGCTCTTGGCGGAAATGAAGAGGATTTCTTACGCAACCTTGATCTGAGATA 840  
QY 841 GCGATCCTTGTAAATGGCGGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
Db 841 GCGATCCTTGTAAATGGCGGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
QY 901 CGCTTGTCTAATAGATCGCAGCCCTTAATTTACCGGTAGTTGAAATGAAGGTGATCTTCT 960  
Db 901 CGCTTGTCTAATAGATCGCAGCCCTTAATTTACCGGTAGTTGAAATGAAGGTGATCTTCT 960  
QY 961 CTTTACGCTTTGGGTAAAGGCTGAGACGGGTGCGCACTCGAAGAAATTTGA 1014  
Db 961 CTTTACGCTTTGGGTAAAGGCTGAGACGGGTGCGCACTCGAAGAAATTTGA 1014

RESULT 3  
ADH36219  
ID ADH36219 standard; DNA; 1014 BP.  
XX

AC ADH36219;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase gene sequence SeqID385.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
XX high throughput system; gene; ds.  
XX Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI McQuaid J, Siege J;  
XX  
DR WPI; 2004-142708/14.  
DR P-PSDB; ADH36220.  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 74; SEQ ID NO 385; 277bp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a gene which encodes a nitrilase enzyme which can be used in the  
CC method of the invention.  
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGCAGACCGCCCCGATCTACATGGATTGGAG 60  
Db 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGCAGACCGCCCCGATCTACATGGATTGGAG 60  
QY 61 GCGACGGTGCACAAAACCATTTAGTTGATGGAAGAAGCAGCAGCTAATAATGCTGCTG 120  
Db 61 GCGACGGTGCACAAAACCATTTAGTTGATGGAAGAAGCAGCAGCTAATAATGCTGCTG 120  
QY 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGTTCTTGGCTTGACTCACCA 180  
Db 121 ATCGCCTTCCGAAACTTGATTCAGGCTACCCATGTTCTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAATCATTTGGAGTTGGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAATCATTTGGAGTTGGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAAGATGCAGCCAAAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
Db 241 CAAGCTAAGCGCATTTCAAGATGCAGCCAAAGCGGTTGGAAATCATGTCACCCCTGGGATG 300  
QY 301 AGTGAACGGGTCCGTGGCACCCTTTACATCAGTCAGTGTTCATAGGCGATATGATGAC 360  
Db 301 AGTGAACGGGTCCGTGGCACCCTTTACATCAGTCAGTGTTCATAGGCGATATGATGAC 360

Db 301 AGTGAACGGGTGGGTGGCACCCTTTACATCAGTCAGTGGTTCATAGGCCAATAATGTGAC 360  
QY 361 ACCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTGTGTGAACGTACTTTGTTCGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTGTGTGAACGTACTTTGTTCGGCGAA 420  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGAGCTGTGTGGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGAGCTGTGTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAAGAGATT 540  
Db 481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAAGAGATT 540  
QY 541 CATGTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCT 600  
Db 541 CATGTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCT 600  
QY 601 GATGTCAATGTAAGCGCCTCTCGAATCTATGCGGTGAAGGCAATGCTTCGTAAGCG 660  
Db 601 GATGTCAATGTAAGCGCCTCTCGAATCTATGCGGTGAAGGCAATGCTTCGTAAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTGTCTTGGCTGTGTGTGTGACACTACGTAATCATAAGGGCTGATGGTGAATTG 780  
Db 721 GCGTGTCTTGGCTGTGTGTGTGACACTACGTAATCATAAGGGCTGATGGTGAATTG 780  
QY 781 GTGCGCGCTCTTGGCGAAATGAAGAGGGTATTCTCTACGCCAAACCTTGATCCTGGAATA 840  
Db 781 GTGCGCGCTCTTGGCGAAATGAAGAGGGTATTCTCTACGCCAAACCTTGATCCTGGAATA 840  
QY 841 CGCATCCTTGTCTAAATGCGCGCAGACCCCTGTGCTCATTTATCCCGTCCCGACATTACT 900  
Db 841 CGCATCCTTGTCTAAATGCGCGCAGACCCCTGTGCTCATTTATCCCGTCCCGACATTACT 900  
QY 901 CGCTTCTAATAGATGCGAGCCCTTAATTAACCGGTAGTTGAAATTGAAGGTGATCTTCGT 960  
Db 901 CGCTTCTAATAGATGCGAGCCCTTAATTAACCGGTAGTTGAAATTGAAGGTGATCTTCGT 960  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGTA 1014  
Db 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGTA 1014  
RESULT 4  
ADG93918 standard; DNA; 1014 BP.  
ID ADG93918 standard; DNA; 1014 BP.  
XX ADG93918;  
AC ADG93918;  
XX 11-MAR-2004 (first entry)  
DT 11-MAR-2004 (first entry)  
XX Nitrilase enzyme gene sequence SeqID385.  
DE Nitrilase enzyme gene sequence SeqID385.  
XX nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
XX enantioselective transformation; gene; ds.  
OS Unidentified.  
XX WO2003097810-A2.  
PN WO2003097810-A2.  
XX 27-NOV-2003.  
PD 27-NOV-2003.  
XX 15-MAY-2003; 2003WO-US015712.  
PF 15-MAY-2003; 2003WO-US015712.  
XX 15-MAY-2002; 2002US-00146772.  
PR 15-MAY-2002; 2002US-00146772.  
XX 09-SEP-2002; 2002US-00241742.  
PA 09-SEP-2002; 2002US-00241742.  
XX (DIVE-) DIVERSA CORP.  
XX (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX WPI; 2004-090637/09.  
DR P-PSDB; ADG93919.  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX Claim 44; SEQ ID NO 385; 295bp; English.  
PS This invention is related to a novel isolated or recombinant nucleic acid  
XX encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAAGCTATCAAGTGCCTGCGTGAAGCCGCCGATCTACATGATTTGGAG 60  
Db 1 ATGAAAGAAGCTATCAAGTGCCTGCGTGAAGCCGCCGATCTACATGATTTGAAG 60  
QY 61 GCGACGGTGGACAAACCATTGAGTTGATGGAAGAAGCAGCAGCAGCTAATATGCTCGTCTG 120  
Db 61 GCGACGGTGGACAAACCATTGAGTTGATGGAAGAAGCAGCAGCAGCTAATATGCTCGTCTG 120  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTAACCCATGTTCTTTGGCTTGAATCACC 180  
Db 121 ATGCGCTTCCGGAACCTTGATTCAGGCTAACCCATGTTCTTTGGCTTGAATCACC 180  
QY 181 GCATGGGCAATGCAATTTGTATGCGCCCAATACGAGAACTCATTTGAGTTGGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTATGCGCCCAATACGAGAACTCATTTGAGTTGGATGGCCCT 240  
QY 241 CAAAGTAAGCGCAATTTGATGCGCCCAATACGAGAACTCATTTGAGTTGGATGGCCCT 300  
Db 241 CAAAGTAAGCGCAATTTGATGCGCCCAATACGAGAACTCATTTGAGTTGGATGGCCCT 300  
QY 301 AGTGAACGGGTGCGTGGCACCCTTACATCAGTCAGTGGTTCATAGCGGATATGTTGAC 360  
Db 301 AGTGAACGGGTGCGTGGCACCCTTACATCAGTCAGTGGTTCATAGCGGATATGTTGAC 360  
QY 361 ACCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTTGTGAACGTAATTTGTTGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTTGTGAACGTAATTTGTTGGCGAA 420  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGAGCTGTGTGGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGAGCTGTGTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAAGAGATT 540  
Db 481 TGTGGAGACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAAGAGATT 540  
QY 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCT 600  
Db 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCT 600  
QY 601 GATGTCAATGTAAGCGCCTCTCGAATCTATGCGGTGAAGGCAATGCTTCGTAAGCG 660  
Db 601 GATGTCAATGTAAGCGCCTCTCGAATCTATGCGGTGAAGGCAATGCTTCGTAAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720



QY	721	GCCTTGCTTCTGCGCTGGTGTGACACTACAGTATCATAGGGCCCTGATGTGTGACTTG	780
Db	721	GCCTTGCTTCTGCGCTGGTGTGACACTACAGTATCATAGGGCCCTGATGTGTGACTTG	780
QY	781	GTCGCGCCTCTTGCCGAAAATGAAGAGGTAATTCTACGCAAACTTGATCCTGAGTA	840
Db	781	GTCGCGCCTCTTGCCGAAAATGAAGAGGTAATTCTACGCAAACTTGATCCTGAGTA	840
QY	841	CGCATCCTTGCTAAATGCGCGCAGACCCCTGCTGTCATTATTCGCCGTCGACATTA	900
Db	841	CGCATCCTTGCTAAATGCGCGCAGACCCCTGCTGTCATTATTCGCCGTCGACATTA	900
QY	901	CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATTGAAGGTGATCTTC	960
Db	901	CGCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATTGAAGGTGATCTTC	960
QY	961	CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGCA	1014
Db	961	CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGCA	1014

CC processes, and "multi-pot" processes using a variety of parameters.  
 CC Atorvastatin is used, in conjunction with dietary restriction, in the  
 CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
 CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
 CC sequence encodes a nitrilase polypeptide obtained from an environmental  
 CC sample.  
 XX  
 SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;



QY 901 CGTTGCTAATAGATCGCAGCCCTAAATTACCGTAGTGAATTGAAGTGATCTTCGT 960  
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db 901 CGCTTGCTAATAGATCGCAGCCCTAAATTACCGTAGTGAATTGAAGTGATCTTCGT 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014  
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
Db 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014  
RESULT 6  
AD164638  
ID AD164638 standard; DNA; 1014 BP.  
XX  
AC AD164638;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE DNA encoding nitrilase seq id 193.  
XX  
KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KW carboxylic acid; cyanhydrin moiety hydrolysis;  
KW aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KW chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KW (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KW (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KW & diastomeric excess; food additive; drug intermediate; ds; nitrilase;  
KW gene.  
XX  
OS Unidentified.  
XX  
PN US2004014195-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 15-MAY-2003; 2003US-00440523.  
XX  
PR 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-121569/12.  
DR P-PSDB; AD164639.  
XX  
PT Novel isolated or recombinant polypeptide having nitrilase activity,  
PT useful in production of food additives.  
XX  
PS Claim 1; SEQ ID NO 385; 105bp; English.  
XX  
CC The invention describes an isolated or recombinant polypeptide (I)  
CC comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (SI) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its  
CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanhydrin moiety or an aminonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining & enantiomeric excess or & diastomeric  
CC excess. (I) is useful in the production of food additives and drug

CC intermediates. This sequence encodes a nitrilase of the invention.  
XX  
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAAGCTATCAAGGTGCGCTGCGCAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 ATGAAAGAAGCTATCAAGGTGCGCTGCGCAAGCCGCCGATCTACATGATTGGAAG 60  
QY 61 GCGACGGTGGACAAAACCATTTAGTTGATGGAAGAACGACGACGTAATAATGCTCGCTG 120  
Db 61 GCGACGGTGGACAAAACCATTTAGTTGATGGAAGAACGACGACGTAATAATGCTCGCTG 120  
QY 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
Db 121 ATCGCCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAAGCTCATTTGGATTGGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAAGCTCATTTGGATTGGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTTCAGATGCAAGCCGATGGGAATCATGTGTCAACCTGGGATG 300  
Db 241 CAAGCTAAGCGCATTTTCAGATGCAAGCCGATGGGAATCATGTGTCAACCTGGGATG 300  
QY 301 AGTGAACGGGTGGTGCGACCCCTTACATCAGTCAAGTTCATAGGCGCATATGTGTAC 360  
Db 301 AGTGAACGGGTGGTGCGACCCCTTACATCAGTCAAGTTCATAGGCGCATATGTGTAC 360  
QY 361 ACCATTGGGGCCCGCGGAAAGTTGAAACCTATTGTTGAAACGTAATTGTTCCGCCGA 420  
Db 361 ACCATTGGGGCCCGCGGAAAGTTGAAACCTATTGTTGAAACGTAATTGTTCCGCCGA 420  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCAGACGTCGTGGAAGGCTGGTGCTTATGC 480  
Db 421 GGGGATGTTTCATCGCTAGCGGTTTTCAGACGTCGTGGAAGGCTGGTGCTTATGC 480  
QY 481 TGTGGAGACACTTCAACCGCTAACAAATAACGCTTGTATGACAAATAAGAGATT 540  
Db 481 TGTGGAGACACTTCAACCGCTAACAAATAACGCTTGTATGACAAATAAGAGATT 540  
QY 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTAATCCTAATGCGGAAAGCCCTGGGGCCT 600  
Db 541 CATTTGCGGCTTGGCCGAGCTTTAGCCTTAATCCTAATGCGGAAAGCCCTGGGGCCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTGACTAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTGACTAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTTGCTTCTGGCTGGTGGTGAGACACTACGTAATCATTAGGGCCTGATGGTGAATTG 780  
Db 721 GCGTTGCTTCTGGCTGGTGGTGAGACACTACGTAATCATTAGGGCCTGATGGTGAATTG 780  
QY 781 GTCCGGCCTTTGCCGAAATGAAGAGGATTTCTTACGCAAACTTGATCTGAGTA 840  
Db 781 GTCCGGCCTTTGCCGAAATGAAGAGGATTTCTTACGCAAACTTGATCTGAGTA 840  
QY 841 CGCATCTTGTCTAATATGGCGGACAGACCTGTGTCATATTCGCGTCCGACATAACT 900  
Db 841 CGCATCTTGTCTAATATGGCGGACAGACCTGTGTCATATTCGCGTCCGACATAACT 900  
QY 901 CGCTTGCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAATGAAGGTGATCTTCGT 960  
Db 901 CGCTTGCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAATGAAGGTGATCTTCGT 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014

Db 961 CCTTACGCTTTGGGTAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA 1014

RESULT 7  
ADCC23780  
ID ADCC23780 standard; DNA; 1014 BP.

AC ADCC23780;  
XX  
DT 18-DEC-2003 (first entry)  
XX

DE DNA sequence (SeqID 47) encoding a nitrilase enzyme.

XX gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.

XX Unidentified.

PN WO2003000840-A2.

XX 03-JAN-2003.

PF 15-MAY-2002; 2002WO-US015983.

XX 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309006P.

PR 22-JAN-2002; 2002US-0351336P.

XX (DIVE-) DIVERSA CORP.

PA (MADD/) MADDEN D.

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,  
PI Short JM, Burk M,  
XX

DR WPI; 2003-201417/19.

DR P-PSDB; ADCC23781.

XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.

XX Claim 1; SEQ ID NO 47; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.

XX Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 10; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 1.1e-152;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGCTCGCTGCGCAAGCGCCCGATCTACATGATTGGAG 61  
DB 2 TGAAGAAGCAATCAAGTAGCCTGTGTGCAAGCACTCCAGCTTTCTCGACTGGAG 61

QY 62 CGACGGTGACAAAACATTGAGTTGATGAGAGACAGCAGTAAATATGCTGCTGA 121  
DB 62 CCACAGTGAGACAGACCGTCGCCCTGATTGAGGAGGACAGCCCGTAAACGGCGCCTTAA 121

QY 122 TCGCCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 181  
DB 122 TCGCCTTCCAGAGACCTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCACCA 181

QY 182 CATGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTC 241  
DB 182 CTTGGGGAGATGCAATTCGTCCGCCATACCAAGAACTCACTGTTCTCGACAGCCCTC 241

QY 242 AAGCTAAGCGCATTTCAAGTAGACCCCAAGCGGTTGGAAATCATGTGACCCCTGGGATGA 301  
DB 242 AGGCCAAGCGCATCAAGTAGAGCCGCCAGCGCCGGTATATACGTCGCGCTAGGGTACA 301

QY 302 GTGAACGGGTGCGTGACCCCTTACATCAGTCAGTGGTTGATAGCGGATATGTTGACA 361  
DB 302 GCGAAGCGGTGAGCGGAACCTCTACATGGGCGAGTGGCTCATTTAGCGAAGGCGGAAA 361

QY 362 CCATTGGGGCCCGCGGAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAAG 421  
DB 362 CAGCTGGGCTGCGCGGAAGCTGAAACCAACCATGTAGAGCGAACCTCTTCGTTGAG 421

QY 422 GGGATGTTTCATCGCTAGCGGTTTTCAGACGCTGTGTTGAAGCGTGGGCTTATGCT 481  
DB 422 GCGACGGATCATCCCTTTCACATTTCAGACACACCGTTGGGGGTGCTGGGCGGACTCTGCT 481

QY 482 GTTGGGAGCACCTTCAACCGCTTAACAAATACGCTTTGATGCACAAATGAAGAGATT 541  
DB 482 GTTGGGAACACTTACAACTCTTTGAAATATGCGCTCTACGACAGAACGAGAAATAC 541

QY 542 ATTGCGGCTTGCGCGAAGCTTACGCTTATCTAATGCGGCGAAGCCCTGGGCGCTG 601  
DB 542 ACTTGGCGGCTTGCGCGAAGCTTACGCTTATCTAATGCGGCGAAGCCCTGGGCGCTG 601

QY 602 ATGTCAATGATGCGGCTCTCGAATCTATGCGGTTGAAGGCAATGCTTGTACTAGCGT 661  
DB 602 AAGTAATGTGCGAGCTTCTCGATCTACGCGGTTGAAGGCAATGCTTGTACTAGCGT 661

QY 662 CGTGTGCGCTGTTTCACAATCCATGATGATGATGCTTTGTACAGATGACGAAAGCATG 721  
DB 662 CCTGCGGCTGCTGCTGCGCAAGATGATGATGATGCTTTGTACAGATGACGAAAGCATG 721

QY 722 CGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
DB 722 GCCTTCTTCAAGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781

QY 782 TCGCGCTCTTTCGCGAAGTGAAGAGGATTTCTTACGCAAACTTGATCTGAGTAC 841  
DB 782 CGCGCGCTTTCGCGAAGTGAAGAGGATTTCTTACGCAAACTTGATCTGAGTAC 841

QY 842 GCATCCTTGTAAATGCGCGCAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901  
DB 842 GAATCTATGCAAGACCGCAGCTGATCCAGCGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCT 901

QY 902 GCTTCTAATAGATGCGACGCCCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 961  
DB 902 GGCTGCTGATCAATGCGACGCCCTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 961

QY 962 CTTACGCTTTGGTAAAGCGCTGAGACCGGTCGCAACTCGAAGAAATTGA 1014  
DB 962 CATCGGCCAAGGCTTTGAAGTTGAGGCGCGCGGCTGAGGAGGCGATTGA 1014

RESULT 8  
ADH35881  
ID ADH35881 standard; DNA; 1014 BP.

XX ADH35881;  
AC  
XX  
DT 11-MAR-2004 (first entry)  
XX

DE Chemical process monitoring-related nitrilase gene sequence SeqID47.  
XX chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; gene; ds.

XX Unidentified.  
OS  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI Mcquaid J, Stege J;  
XX  
XX WPI; 2004-142708/14.  
DR P-PSDB; ADH35882.  
XX  
XX Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 74; SEQ ID NO 47; 277pp; English.  
XX  
XX This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a gene which encodes a nitrilase enzyme which can be used in the  
CC method of the invention.  
XX  
SQ Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 12; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 1.1e-152;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGATCTACATGATTGGAGG 61  
Db |||||  
2 TGAAGAAGCAATCAAGTAGCCTGTGTGCAAGCAGCTCTTCTCGACCTGGACG 61  
QY 62 CGACGGTGAACAAAACATGAGTTGATGGAAGAAGCAGCAGTAATAATGCTCGTCTGA 121  
Db |||||  
62 CCACAGTGAACAAGACCGTCGCTGATTGAGAGGAGCGCCCGTAACGCGCAGCCTAA 121  
QY 122 TCGCCTTCCGGAACCTTGATCCAGGCTACCCATGGTTCTTGGCTGACTCAGCAG 181  
Db |||||  
122 TCGCCTTCCAGAGACCTGATCCAGGCTACCCATGGTTCTTGGCTGACTCAGCAG 181  
QY 182 CATGGGCAATGCAATTTGATCGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241  
Db |||||  
182 CTTGGGGGATGCAATTCGTGCGCGGATACCAAGAACTCATCTGCTCGACAGCCCTC 241  
QY 242 AAGCTAAGCGCATTTGATGATGAGCCCAAGCGGTTGGGAATCATGTCACCCCTGGGGATGA 301  
Db |||||  
242 AGGCCAAGCGCATCATGAGGCGCGCCCAAGCGCGCGGTATATACGTGCGGCTAGGGGTACA 301  
QY 302 GTGAACGGGTGCGGTGGCACCCCTTACATCATGTCAGTGGTTCTATAGGCGATATGTTGATGACA 361  
Db |||||  
302 GCGAAGCGGTGAGCGGAACCCCTTACATGAGGCGAGTGGCTCATTTGACGATTAAGGCGGAAA 361  
QY 362 CCATTGGGGCCCGCGGAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTGGGCGAAG 421  
Db |||||  
362 CAGCTGGGCTGCGCGGAAGCTGAACCAACCATGTAGAGCGAACCCCTCTTGGGTGAAG 421  
QY 422 GGGATGGTTTCATCGCTAGCGGTTTTCGAGACGTCTGTTGGAAGGCTGGGCTTATGCT 481  
Db |||||

Db 422 GCGAGGATCATCCCTTCCACTTTCGACACACCGTTGGGGTGCTGGCGGACTTGCT 481  
QY 482 GTTGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAAATGAAGAGATTG 541  
Db |||||  
482 GTTGGGAACACTTACAACTCTTTCGAATATGCGCTCAGCAGACAGAGAAATAC 541  
QY 542 ATTTGGGCGCTTGGCGGAGCTTTAGCCTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601  
Db |||||  
542 ACTTGGCGGCTTGGCGGAGCTTACGATCTACCGTCAAGCGACAGAAAGTCTTGACCAAG 601  
QY 602 ATGTCAATGTAGCGGCGCTTCGAAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCGT 661  
Db |||||  
602 AAGTAATGTGCGACGCTTCTCGGATCTACGCCGTGAAGGGCAGTGTGTTGTTCTCGCTT 661  
QY 662 CGTGTGCGCTGCTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAAGCATG 721  
Db |||||  
662 CCTGCGGCTGCTCTCGCCAGAGATGATCGAAATGCTCTGCACTGACGAAAAGCAAGCACA 721  
QY 722 CGTTGCTTCTGCGCTGCTGTTGACACTCAGTATCATAGGGGCTGATGGTGTGACTTGG 781  
Db |||||  
722 GCCTTCTTCAAGCGCGGCGGCGGCTACTCCGCAATATCGGTGCCATGGCAGCAGCCTAG 781  
QY 782 TCGCGCTCTTGGCGAAATGAAGGATATTCTTACGCAAACTTGATCCTGGAGTAC 841  
Db |||||  
782 CGCGCCCTTGGGCGAAAACGAGAAAGTATTCTATGACCACCTGAGCCCTGCGGCTC 841  
QY 842 GCATCCTTGTAAATGGCGGACAGACCTGCTGTCATATTATCCGTCGACATTACTC 901  
Db |||||  
842 GAATCTATGCAAAAGACCGCAGCTGATCCAGCGGGCACTACTCCAGACCAAGCAGTCACTC 901  
QY 902 GCTTGTATATGATCGCAGCCCTAAATACCGGTAGTTGAAATTGAAGGTGATCTTCTGTC 961  
Db |||||  
902 GGCTGTGATCAATCGCAGTGCACATCAGCCAGTCGTAGAGTTGGAAGGAAATACCTG 961  
QY 962 CTTACGCTTTGGGTAAGCGTCTGAGACGGGTGCCAATCGAAGAAATTGA 1014  
Db |||||  
962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCGCGGCTGACGAAGCGCAATTGA 1014

RESULT 9  
ADG93582  
ID ADG93582 standard; DNA; 1014 BP.

ADG93582;  
11-MAR-2004 (first entry)

Nitrilase enzyme gene sequence SeqID47.

nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
enantioselective transformation; gene; ds.

Unidentified.

WO2003097810-A2.

27-NOV-2003.

15-MAY-2003; 2003WO-US015712.

15-MAY-2002; 2002US-00146772.

09-SEP-2002; 2002US-00241742.

(DIVE-) DIVERSA CORP.

Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

WPI; 2004-090637/09.

P-PSDB; ADG93583.

New isolated or recombinant nucleic acid encoding a polypeptide having  
nitrilase activity, useful for screening enantioselective transformation.







Query Match		49.3%;	Score 499.4;	DB 12;	Length 1014;
Best Local Similarity		68.3%;	Pred. No. 1.1e-152;		
Matches 692;		Conservative 0;	Mismatches 321;	Indels 0;	Gaps 0;
QY	2	TGAAAAGAACTATCAAGGTCGCTGCGTGCAGAGCCCGCCGATCTACATGATTGGAGG	61		
Db	2	TGAAAAGAAACAATCAAGTAGCTGTGTGCAAGCAGCTCCAGTCTTCTCGACCTGGACG	61		
QY	62	CGACGGTGGACAAAACCATTTGATTTGATGGAAGAGCAGCAGTAATAATGCTCGTCTGA	121		
Db	62	CGACAGTGGACAAGACCGTCGCCCTGATTGAGGAGCAGCCGTAACGGCGCAGCCTAA	121		
QY	122	TGCGCTTTCGGAAACTTGGATTCCAGGCTAACCCATGTTTCTTTGGCTTGACTCACCAG	181		
Db	122	TGCGCTTTCAGAGACCTGATTCAGGCTAACCCATGTTTCTTTGGCTTGACTCACCAG	181		
QY	182	CATGGGCAATGCAATTTGTACGCCAATACCATGAGAAGTCAATGAGTTGGATGCCCTC	241		
Db	182	CTGGGGGATGCAATTGTCGCCCGGATACCAAGAACTCACTGTCCTCGACAGCCCTC	241		
QY	242	AAGCTAAGCGCATTTCAAGATGCAGCCAAAGCGTTGGGAATCATGTCACCCGCGGATGA	301		
Db	242	AGGCCAAGCGCATCAGTGAGGCCGCCCAAGCGCGCGGTATATACGTGCCCTAGGGTACA	301		
QY	302	GTGAACGGGTCGGTGGCACCCCTTTACATCACTCACTGTTGTTGTAAGCGGATATGTTGACA	361		
Db	302	GCGAACGCGTGAAGCGGAACCTCTACATGGGGCAGTGGCTCATTTGACGATTAAGGGCGAAA	361		
QY	362	CCATTGGGGCGCGCGGAAAGTTGAAACCTACTTTTGTGAACGTAATTTGTTGGCGAAG	421		
Db	362	CAGCTGGGCTGCGCGCGAAAGCTGAAACCAACCATGTAGAGCGAAACCTCTTCGGTGAAG	421		
QY	422	GGATGGTTCAATCGCTAGCGGTTTTCAGACGTCTGTTGAAAGGCTGGGTGGCTTATGCT	481		
Db	422	GCGACGGAATCACTCCCTTTCACACTTTCAGACACCGTTGGGGTGTGGCGGACTCTGCT	481		
QY	482	GTTGGAGACCTTCAACCGCTAACAAAATACGCTTTGTATGACAAAATGAAGAGATTG	541		
Db	482	GTTGGAGACACTTACAACTCTTTCGAAATATGCGCTCTACGACAGAAAGAAATATAC	541		
QY	542	ATTGTGCGGCTTGGCCGAGCTTTAGCCTTTATCTAATGCGGGAAGCCCTGGGGCCTG	601		
Db	542	ACTTGGCCGCTGCTGCTTAGCTTACGATCTTACCGTCAAGCGACAGAAAGTCTTGGACCA	601		
QY	602	ATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAAGGGCAATGCTTCGTAAGCGT	661		
Db	602	AAGTAATGTGCACTTCTCGATCTACGCCGTGGAAGGGCAGTGTCTTCTCGCTT	661		
QY	662	CGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAAGCATG	721		
Db	662	CCTGCGGCTCGTCTGCGCAGAGATGATGAAATGCTCTGCACTGACGAAAGCAAGCACA	721		
QY	722	CGTGTCTTGTGCTGTGTGGACACTCACGTAATCATAAGGGCTGATGTTGTGACTTGG	781		
Db	722	GCCTTCTTCAAGCGCGCGGGGTACTCCCGCATTAATCGGTCCGATGGCAGCGACCTAG	781		
QY	782	TGCGGCTCTTGGCCGAAATGAAGAGGTATTCTTAACGCAAACTTGATCTGAGATAC	841		
Db	782	CGCGCCCTTGGGCGGAAAAGAGAGATATTCTTAATGCCACTTGACCCGCGCTC	841		
QY	842	GCATCCTTGTAAATGGCGGACAGACCCCTGCTGTGTAATTTCCCGTCCCGACATACTC	901		
Db	842	GAACTATGCAAAAGACCGAGCTGATCCAGCCGGGCACTACTCCAGACCAAGACGTCACTC	901		
QY	902	GCTTGTATATAGATCGCAGCCCTAAATTACCGGTAGTTGAAATGGAAGGTGATCTTCGTC	961		
Db	902	GGCTGTGATCAATCGCAGTGCATATCAGCCAGTCGTAGAGGTTGAAAGGGAATATACCTG	961		
QY	962	CTTACGCTTTGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA	1014		
Db	962	CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCCGCCCGGTACGAAAGCGGATTGA	1014		

RESULT 11	
AD164300	AD164300 standard; DNA; 1014 BP.
ID	AD164300
XX	
AC	AD164300;
XX	
DT	22-APR-2004 (first entry)
XX	
DE	DNA encoding nitrilase seq id 24.
XX	
KW	(R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
KW	carboxylic acid; cyanohydrin moiety hydrolysis;
KW	aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
KW	chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
KW	(R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
KW	(R)-phenyl lactic acid derivative; & enantiomeric excess;
KW	& diastereomeric excess; food additive; drug intermediate; ds; nitrilase;
KW	gene.
XX	
OS	Unidentified.
XX	
PN	US2004014195-A1.
XX	
PD	22-JAN-2004.
XX	
XX	15-MAY-2003; 2003US-00440523.
PF	
XX	29-DEC-1999; 99US-0173609P.
PR	07-DEC-2000; 2000US-0254414P.
PR	28-DEC-2000; 2000US-00751299.
PR	21-JUN-2001; 2001US-0300189P.
PR	30-JUL-2001; 2001US-0309006P.
PR	22-JAN-2002; 2002US-0351336P.
PR	15-MAY-2002; 2002US-00146772.
PR	09-SEP-2002; 2002US-00241742.
XX	
PA	(DIVE-) DIVERSA CORP.
XX	
PI	Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX	
XX	WPI; 2004-121569/12.
DR	P-PSDB; AD164301.
XX	
PT	Novel isolated or recombinant polypeptide having nitrilase activity,
XX	useful in production of food additives.
XX	
PS	Claim 1; SEQ ID NO 47; 105pp; English.
XX	
CC	The invention describes an isolated or recombinant polypeptide (I)
CC	comprising amino acids having a sequence at least 50 % identical to a
CC	sequence (S1) available in electronic form (EC) from the following web
CC	site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its
CC	variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC	residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC	fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC	hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC	hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
CC	chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC	producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC	mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
CC	derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
CC	and for identifying a modified compound. The inventive method is useful
CC	for monitoring or determining & enantiomeric excess or & diastereomeric
CC	excess. (I) is useful in the production of food additives and drug
CC	intermediates. This sequence encodes a nitrilase of the invention.
XX	
SQ	Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match		49.3%;	Score 499.4;	DB 12;	Length 1014;
Best Local Similarity		68.3%;	Pred. No. 1.1e-152;		
Matches 692;		Conservative 0;	Mismatches 321;	Indels 0;	Gaps 0;

QY	2	TGAAAGAGACTATCAAGGTCGCTGCGTGCAGCGCCCGCATCTACATGGATTGGAGG	61
Db	2	TGAAAGAGCAATCAAAAGTAGCCTGTGTGCAAGCACTCCAGTCTTCTCGACCTGGACG	61
QY	62	CGACGGTGGACAAAAACCATTTGATGTGATGGAAGACGACGCTAATATAGCTGCTGA	121
Db	62	CCACAGTGGACAAAGACCGTCGCCCTGATTGAGGAGGCAGCCCGTAAACGCGCAGCCTAA	121
QY	122	TCGCTTTCCGAAAACCTTGATTCAGGCTACCGTAGTGTCTTTGGCTTGACTCACCAG	181
Db	122	TCGCTTTCCAGAGACCTGGATTCAGGCTACCGTAGTGTCTTTGGCTGACTCACCAG	181
QY	182	CATGGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC	241
Db	182	CCTGGGGGATGCATTTGCTGCGCCGATACCAAGAACTCACTGGTCTCGACAGCCCTC	241
QY	242	AAGCTAAGCCCATTTTCAGATGCAGCCAAAGCGGTGGGAATCATGCTCACCTGGGGATGA	301
Db	242	AGGCCAAGCCCATTCAGTAGAGCCGCCAGCGCGCGGTATATACGTGCGCTAGGGTACA	301
QY	302	GTTGAACGGGTCGGTGGCACCCCTTACATCAGTCAAGTTCATAGCGGATAATGTGACA	361
Db	302	GCGAACCGCTGAGCGGAACCTCTACATGGGGCAGTGGCTCATTTAGCATAAAGGCGAAA	361
QY	362	CCATTGGGGCCCGCGGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTGGCGGAAG	421
Db	362	CAGCTGGGCTGCGCGGAAAGCTGAAACCAACCATGTAGACGAACCCCTCTTGGTGAAG	421
QY	422	GGGATGTTTCATCGCTAGCGGTTTTTCGAGACGTCTGTGGAAGGCTGGGTGCTTATGCT	481
Db	422	GCGACGATCATCCCTTCCACTTTCGACACACCGTTGGGGGTGCTGGCGGACTCTGCT	481
QY	482	GTTGGGAGCACCTTCAACCGCTAAACAAAATACGCTTTGTATGCACAAAATGAAGATTC	541
Db	482	GTTGGGAACACTTACAACTCTTTCGAAATATGGCTCTACGACAGAACGAGGAATAC	541
QY	542	ATTGTGCGGCTTGCGCGAGCTTTAGCCTTATCTAATCCGGCGAAAGCCCTGGGGCTG	601
Db	542	ACTTCCGCCCTGCGCTAGCTTCAAGCATTAACGTCACAGGACAGAAAGTCTTGGACACG	601
QY	602	ATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCGTACTAGCGT	661
Db	602	AAGTAATGTCCGAGCTTCTCGATCTACGCGGTGGAAGGGCAGTGTGTTCTCGCTT	661
QY	662	CGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATG	721
Db	662	CTGCGCGCTCGTCTCGCAGAGATGATCGAAATGCTCTGCACTGACGAAAGCAAGACA	721
QY	722	CGTTGCTTCTGGCTGTGTGACACTCAAGTATCATAGGGCGCTGATGGTGTGACTTGG	781
Db	722	GCTTCTTCAAGCGCGCGGCTACTCCGCATTAATCGGTCCCGATGGCAGCAGCTAG	781
QY	782	TCGCGCTCTTGCCGAAAATGAAGGGTATTCTTACGCAAACTTGATCTTGAGATAC	841
Db	782	CGCGCCCTTGCGCGGAAAAAGAGAAAGTATTCTATGCCACTGTGACCCCTGCGCTC	841
QY	842	GCATCCTTGCTAAATGCGCGCAGACCTTGCTGTCATTATTCCCGTCCCGACATTAATC	901
Db	842	GATCTATGCAAAAGACCGCAGCTGATCCAGCCGGGCACTACTCCAGACCAAGACGTCACTC	901
QY	902	GCTTGCTAATAGATCGCAGCCCTTAATTAACCGGTAGTTGAAATGAAAGGTGATCTTCGTC	961
Db	902	GGCTGCTGATCAATCGCAGTGCCAATCAGCCAGTCTAGAGGTTGGAAGGGAATACCTG	961
QY	962	CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGGA	1014
Db	962	CATCGGCCCAAGGCTTTGAAAGTTGAGCGCGCCCCCGGTACGAAAGCGGATTTGA	1014

RESULT 12	
ADC23790	
ID	ADC23790 standard; DNA; 1014 BP.
XX	

AC		ADC23790;	
XX	DT	18-DEC-2003	(first entry)
XX	DE	DNA sequence (SeqID 57) encoding a nitrilase enzyme.	
XX	KW	gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst; enantiomer; chiral medicine.	
OS		Unidentified.	
PN		WO2003000840-A2.	
PD		03-JAN-2003.	
PX		15-MAY-2002; 2002WO-US015983.	
PR		21-JUN-2001; 2001US-0300189P. 30-JUL-2001; 2001US-0309006P. 22-JAN-2002; 2002US-0351336P.	
PA		(DIVE-) DIVERSA CORP. (MADD/) MADDEN D.	
PI		Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E, Short JM, Burk M;	
PX		WPt; 2003-201417/19.	
DR		P-PSTDB; ADC23791.	
XV		Noval nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano- 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.	
PT		Claim 1; SEQ ID NO 57; 560pp; English.	
PS		This invention relates to nitrilases and the nucleic acids that encode these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydridins into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophillic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polynucleotide is a DNA sequence that encodes a nitrilase enzyme of the invention.	
CC		Sequence 1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;	
SQ			
Query Match		49.1%; Score 497.8; DB 10; Length 1014;	
Best Local Similarity		68.2%; Pred. No. 3.6e-152;	
Matches 691; Conservative		0; Mismatches 322; Indels 0; Gaps 0	
OY		2 TGAAGAAGCTATCAAGGTCGGCTCGGTGCAGCAGCCCCCGATTACATGATTGGAGG	61
Db			
		2 TGAAGAAGCAATCAAAGTAGCCGTGTGTGCCAACGACTCCAGTCTTTCTGCAGCTGAGC	61
OY		62 CGACGGTGACA AAAACCATTGAGTTGATGAA GAAGCACAGTAATAATGCTGCTGA	121
Db			
		62 CCACAGTGACAAAGACCGTCGCCCTGATGAGAGGACGAGCCCCTAAACGGCGACGCCCTAA	121
OY		122 TC GCCCTTCCGGA ACTTGA ATTCAGGCT ACCCATG GTTCTTTGGCTTGA CTCA CCAG	181
Db			
		122 TC GCCTTCCAG AGACCTG GATTCAGGCT AC CATG GTTCTTTGGCTTGA CTCA CCAG	181
OY		182 CATGGGCAATGCA ATTGTG TAGCCA ATA CCATGAGA AACTCA TTGAGTTGGATGGCCCTC	241
Db			
		182 CCTGGGGGATGCA ATTCTGTG CCGGAT AACGAGAA CTA CTGCTCTGACAGCCCTC	241
		CCTGGGGGATGCA ATTCTGTG CCGGAT AACGAGAA CTA CTGCTCTGACAGCCCTC	241

QY 182 CATGGGCATGCAATTGTACGCCAATACCATGAGAACTCATTTGAATTGGATGGCCCTC 241  
| | | | | | | | | | | | | | | | | | | | |  
Db 182 CCGTGGGGATGCAATTCGTGCGCCGATAACACGAGAAGCTACTGTCTTCGACAGCCCTC 241

QY		242	AAGCTAAGCGCATTTTCAGATGCAGCCGCGGTGGGAATCATTGTCACCTGGGGATGA	301
Db		242	AGGCCAAGCGCATCAGTAGAGCGCGCCAGCGCCGGTATATACGTCCGCTAGGGTACA	301
QY		302	GTCGAACGGGTCCGTGCAACCCTTTACATCAGTCAGTGTTTCATAGCGCAATAAGTGACA	361
Db		302	GCGAACCGGTGAGCGGAACCCCTCACATGGGGCAGTGCGTCATTGACGATAAGGCGAAA	361
QY		362	CCATTGGGGCCCCCGGAAAGTTGAACCTACTTTGTTGAACGTACTTTGTTCCGGGAAG	421
Db		362	CAGCTGGGGCTGCCCGGAAAGCTGAACCAACCATGTAGAGCGAACCCCTTCGGTGAAG	421
QY		422	GGGATGGTTTCATCGCTAGCGGTTTTCCAGA C GTCTGTGGAAGGCTGGGTGCTTATGCT	481
Db		422	GCGACGGATCATCCCTTTCCACTTTCGACACACCGTTGGGGGTGCTGGGCGGACTCTGCT	481
QY		482	GTTGGGAGCACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAGAGATTTC	541
Db		482	GTTGGGAACACTTACCAACCTCTTTCGAATATGCGCTTACGCACAGAACGAGAAATAC	541
QY		542	ATTGTCGAGCTTGCCGAGCTTTAGCCCTTTATTCCTAATGCGGCGAAAGCCCTGGGGCTG	601
Db		542	ACTTCGCGCCTGGCCTTAGCTTCAGCATCTACCGTCAAGCGACAGAAAGTCCTTGGACCA	601
QY		602	ATGTCAATGTAGCGGCTCTCGAATCTATGCCGTGAAGGGCAATGCTTCGTAACGT	661
Db		602	AAGTAATGTGCGAGCTTCTCGATCTAGCGCCGTGAAGGGCAGTGTGTTGTTTCGCTT	661
QY		662	CGTGTGCGCTCGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAACGATG	721
Db		662	CCTGCGCGCTCGTCTCGCAGAGATGATCGAAATGCTCTGCACTGACGAAAGCAGACA	721
QY		722	CGTTGCTTCTGCGCTGTGTGTGACACTCA CGTATCATAGGGCTGATGTTGCTGACTTG	781
Db		722	GCCTTCTTCAAGGCGCGCGGGGTACTCCC GCATTATCGGTCCGATGGCAGCGACTAG	781
QY		782	TGCGCGCTTGTGCGGAAAAATGAAGAGGATTTCTTACGCCAAACCTTGATCCTGGAGTAC	841
Db		782	CGCGCCCTTGGCGGAAAAAGAGAAAGTATTCTTATGCCACTCTGGAACCTGCCGCTC	841
QY		842	GCACTCTTGTSTAAATGCGCGGACAGCCCTGCTGTGTCATTATTCCTCCGACATTA CT	901
Db		842	GAATCTATGCAAGACCGCAGCTGATCCAGCCGGGCACTACTCCAGACGAGCTCACTC	901
QY		902	GCTTGCTAATGATGCGCAGCCCTAAATTA CCGGTAGTTGAAATTGAAGGTGATCTTCGT	961
Db		902	GGCTGCTGATCAATGCGCAGTGC CAATCAGCCAGTCGTAGAGGTGGA CGGAAATACCTG	961
QY		962	CTTAGCTTTGGGTAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA	1014
Db		962	CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCCCCCGGTACGAGGCGATTGA	1014

RESULT 13  
ADH35891 standard; DNA; 1014 BP.

ADH35891;  
11-MAR-2004 (first entry)

Chemical process monitoring-related nitrilase gene sequence SeqID57.

chemical process monitoring; biochemical process monitoring; cyanide;  
high throughput system; gene; ds.

Unidentified.

W02003098187-A2.  
27-NOV-2003.

15-MAY-2003; 2003WO-US015639.

XX	
PR	15-MAY-2002; 2002US-0380737P.
XX	
PA	(DIVE-) DIVERSA CORP.
XX	
PI	Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI	Mcquaid J, Stege J;
XX	
DR	WPI; 2004-142708/14.
DR	P-PSDB; ADH35892.
XX	
PT	Monitoring a chemical or biochemical process comprises providing a
PT	reactant comprising a cyanide or a material that can be converted to
PT	cyanide or a reactant that generates a cyanide or a material that can be
PT	converted to cyanide.
XX	
PS	Claim 74; SEQ ID NO 57; 277pp; English.
XX	
CC	This invention relates to a novel method of monitoring chemical or
CC	biochemical processes. The method involves providing a reactant
CC	comprising cyanide (or a material that can be converted to a cyanide)
CC	that generates as a reaction product cyanide or a material that can be
CC	converted to cyanide and measuring the concentration of produced cyanide
CC	The method is useful for monitoring a chemical or biochemical process.
CC	The method is effective for high throughput systems and is sufficiently
CC	sensitive to detect a small amount of product. The present sequence is
CC	that of a gene which encodes a nitrilase enzyme which can be used in the
CC	method of the invention.
XX	
SO	Sequence 1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;

Query Match	49.1%;	Score 497.8;	DB 12;	Length 1014;
Best Local Similarity	68.2%;	Pred. No. 3.6e-152;		
Matches 691;	Conservative 0;	Mismatches 322;	Indels 0;	Gaps 0;
QY	2	TGAAGAAGCTATCAAGTCGCTGCGTGCAAGCCGCCGATCTACATGATTTGGAGG	61	
Db	2	TGAAGAAGCAATCAAGTAGCTGTGTGCAAGCAGCTCCAGTCTTCTCGACCTGGACG	61	
QY	62	CGACGGTGGACAAAACCATGTAGTTGATGGAAGAAGCAGACGTAAATATGCTGCTGA	121	
Db	62	CCACAGTGGACAAAGACCGTCGCTGATTGAGAGGACGCCGTAAACGGCAGCCTAA	121	
QY	122	TGCGCTTTCGGAACTTGAATCCAGGCTACCCATGTTTCTTGCTTGACTCACAG	181	
Db	122	TGCGCTTTCAGAGACCTGATTCAGGCTACCCATGTTTCTTGCTTGACTCACAG	181	
QY	182	CATGGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGGATGSCCTC	241	
Db	182	CTGGGGGATGCATTCGTGCGCGGATACCAAGAACTCAGTGTCTCGACAGCCTC	241	
QY	242	AAGCTAAGCCGATTTACAGATGACCCAGCGGTTGGGAATCATGCTCACCCCTGGGATGA	301	
Db	242	AGGCCAAGCCGATCAGTAGGCGCCGCCAGCGCGGTATATACGTCGCGCTAGGGTACA	301	
QY	302	GTGAACGGGTCGGTGGCACCCCTTACATCAGTCAGTGTTCATAGCGGATAATGTTGACA	361	
Db	302	GCGAACCGGTGAGCGGAACCTCTACATGCGGAGTGCTCATTTGACGATAAGGGCGAAA	361	
QY	362	CCATTGGGCCCCGGGAAAGTTGAACTACTTTGTGTAACGTACTTTGTTCCGGCAG	421	
Db	362	CAGCTGGGCTGCGCGAAAGCTGAAACCAACCCATGTAGAGCGAACCTCTTCGGTGAAG	421	
QY	422	GGATGGTTCATGCTAGCGGTTTTCAGACGTCGTGTTGAAGGCTGGGCTTATGCT	481	
Db	422	GCGACGATCATCCCTTTCACATTGACACACCGTTGGGGTGTGGGCGGACTCTGCT	481	
QY	482	GTTGGAGACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGAGATTTC	541	
Db	482	GTTGGGAACACTTACAACCTCTTTCGAATATAGCGCTCTACGACAGACGAGAAATAC	541	
QY	542	ATTGTGGGCTTGGCGGAGCTTACCTTATCTATATGCGGGAAGCCCTGGGGCTG	601	



Db 542 ACTTCGGCCGCTGGCCCTAGCTTCAGCATCTACCGTCAAGCGACAGAAGTCTTGACCAg 601  
QY 602 ATGTCAATGTAGCGGCGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCGTACTAGCGT 661  
Db 602 AAGTAAATGTGCAGCTTCTCGGATCTACGCCGTGAAGGCAAGTGTGTTGTTCTCGCTT 661  
QY 662 CGTGTGCGCTCGTTTCAATCCATGATGATATGCTTTGTAACAGATGACGAAAGCATG 721  
Db 662 CTGCGCGCTCGTCTCGCCAGAGATGATCGAATGCTCTGCACTGACGAAAGCAAGACA 721  
QY 722 CGTGTCTTGTGGCTGTGTGTGACACTCAGCTATCATAGGCGCTGATGTGTGACTTGG 781  
Db 722 GCCTTCTTCAAGCGCGCGGGGTACTCCGCATTATCGGTCCCGATGGCAGCGACTAG 781  
QY 782 TCGCGCCTCTTGCCGAAATGAAGGGTATTCTACGCAACCTTGATCCTGAGTAC 841  
Db 782 CGCGCCCTTGGCGGAAAGAGGAGTATTCTATGCCACTCTGACCCCTGCCGCTC 841  
QY 842 GCATCCTTGCTAAATGGCGGACAGCCCTGTGTATATTCGCCGACCATTAATC 901  
Db 842 GAATCTATGCAAGACCGGAGCTGATCCAGCCGGCAGTACTCCAGACGACGTAATC 901  
QY 902 GCTTGCTATAGATCGCAGCCCTAATTAACCGGTAGTTGAATTTGAAGGTGATCTGTC 961  
Db 902 GCCTGCTGATCAATCGCAGTGCCTCAATCAGCCAGTCTGAGAGTTGACGGGAATACCTG 961  
QY 962 CTTAAGCTTTGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTTGA 1014  
Db 962 CATCGGCCCAAGGCTTTGAAGTTGAGCGCGCCCGGGTACGAGCGGATTGA 1014

RESULT 14  
ADG93592  
ID ADG93592 standard; DNA; 1014 BP.

XX ADG93592;

DT 11-MAR-2004 (first entry)

XX Nitrlase enzyme gene sequence SeqID57.

XX nitrlase; nitrlle; carboxylic acid; chemical process; pH; temperature;

KW enantioselective transformation; gene; ds.

XX Unidentified.

PN WO2003097810-A2.

XX 27-NOV-2003.

PF 15-MAY-2003; 2003WO-US015712.

XX 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

XX (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

XX WPI; 2004-090637/09.

DR P-PSDB; ADG93593.

XX New isolated or recombinant nucleic acid encoding a polypeptide having

PT nitrlase activity, useful for screening enantioselective transformation.

XX Claim 44; SEQ ID NO 57; 295bp; English.

XX This invention is related to a novel isolated or recombinant nucleic acid

CC encoding a protein having nitrlase activity. Nitrlase's are capable of

CC converting nitrlle's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated

CC nitrlase proteins of the invention have increased activity and stability

CC at increased pH and temperature when compared to those conventionally

CC used. In addition, the nucleic acid of the invention is useful for

CC screening enantioselective transformation. The present sequence is that

CC of a DNA sequence which encodes a nitrlase enzyme of the invention.

XX Sequence 1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;

SO Query Match 49.1%; Score 497.8; DB 12; Length 1014;

QY Best Local Similarity 68.2%; Pred. No. 3.6e-152;

Db Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGTCCGCTGCGTCAAGCCCGCCGATCTACATGATTTGAGC 61

Db 2 TGAAGAAGCAATCAAGTACCTGTGTGCAAGCAGCTCCAGTCTTCTCGACCTGAGC 61

QY 62 CGACGGTGACAAACCAATGAGTTGATGGAAGAGACGCACTAATATGCTCGTCTGA 121

Db 62 CCAAGTGACAAAGACCGTCCGCTGATTGAGGAGAGCCGTAACGGCGCAGCCTAA 121

QY 122 TCGCCTTCCGAAACTTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCAGCAG 181

Db 122 TCGCCTTCCAGAGACCTGATTCAGGCTACCCATGCTTCTTGGCTTGACTCAGCAG 181

QY 182 CATGGCAATGCAATTTGTAAGCCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241

Db 182 CTGCGGGATGCAATTTGTCGCCCGATACCAAGCACTCACTGCTCCTGACAGCCCTC 241

QY 242 AAGCTAAGCGCATTTGATGACGCCAAGCGGTTGGAAATCATGCTACCCCTGGGATGA 301

Db 242 AGGCCAAGCGCATGATGAGGCCGCCAGCGCGGTTATATACGTCGCGCTAGGGTACA 301

QY 302 GTGAACGGGTGCGGACCCCTTACATCAGTGTGCTCATAGGCGATATAGGTATGAGCA 361

Db 302 GCGAAGCGGTGAGCGGAACCTCTACATGCGGAGTGGGCTCATTTGACGATTAAGCGGAAA 361

QY 362 CCATTGGGCGCGCGGAAAGTTGAAACCTACTTTGTTGAAAGTACTTTGTTGCGCGAAG 421

Db 362 CAGCTGGGCTGCGCGGAAAGCTGAAACCAACCATGTAGAAGCGAACCTCTTGGTGAAG 421

QY 422 GCGATGTTCAATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGGCTTATGCT 481

Db 422 GCGACGATCATCCCTTTCACCTTTCGACACACCGTTGGGGTGCTGGCGGACTCTGCT 481

QY 482 GTTGGAGACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATGAAGAGATTTC 541

Db 482 GTTGGGAACACTTACAACTCTTTGAAATATGCGCTTACGACAGAAAGAGAAATAC 541

QY 542 ATTGTGCGGCTTGGCGGAGCTTTAGCCCTTATCTAATGCGGCGAAAGCCCTGGGGCTG 601

Db 542 ACTTGGCGGCTTGGCTTACCTTACGATCTACCGTCAAGCGACAGAAAGTCTTGACACAG 601

QY 602 ATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGCAATGCTTCGTACTAGCGT 661

Db 602 AAGTAAATGTGCAGCTTCTCGGATCTACGCGGTGAAGGCAAGTGTGTTGTTCTCGCTT 661

QY 662 CGTGTGCGCTCGTTTCAATCCATGATGATATGCTTTGTACAGATGACGAAAGCATG 721

Db 662 CCTGCGGCTCGTCTCGCCAGAGATGATCGAATGCTCTGCACTGACGAAAGCAAGACACA 721

QY 722 CGTGTCTTGTGGCTGTGTGTGACACTCAGCTATCATAGGGCTGATGTGTGACTTGG 781

Db 722 GCCTTCTTCAAGCGCGCGGGTACTCCGCATTATCGGTCCTGATGGCAGCAGCTAG 781

QY 782 TCGCGCCTCTTGGCGGAAATGAAGAGGGTATTCTTACGCAAACTTGATCTGAGATAC 841

Db 782 CGCGCCCTTGGCGGAAAGAGGAGTATTCTTATGCCACTCTGAGACCTGCGGCTC 841

QY 842 GCATCCTTGCTAAATGGCGGACAGCCCTGTGCTGATTAATCCCGTCCGACATTAATC 901

Db 842 GAATCTATGCAAGACCGCAGCTGATCCAGCCGCGCACTACTCAGACGACGATCACTC 901

QY 902 GCTTGCTAATAGATCGAGCCCTAAATTAACCGGTAGTTGAATTTGAAGGTGATCTTCTGTC 961

Db 902 GCCTGCTGATCAATCGCAGTGCACATCAGCCAGTCTGTAAGAGTTGACGCGGAATACTCTG 961





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:54:39 ; Search time 209.215 Seconds  
(without alignments)  
8615.305 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaaagaagctatcaaggt.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/1\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/H\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/ina/PP\_COMB.seq:\*  
8: /cgn2\_6/ptodata/1/ina/RE\_COMB.seq:\*  
9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	283	27.9	1035	2	US-08-690-493-2 Sequence 2, Appli
2	283	27.9	1200	2	US-08-690-493-3 Sequence 3, Appli
3	248.6	24.5	1071	3	US-09-806-876A-1 Sequence 1, Appli
4	218	21.5	1110	3	US-09-823-373-4 Sequence 4, Appli
5	218	21.5	1110	3	US-09-823-373-13 Sequence 13, Appli
6	218	21.5	1176	3	US-09-823-373-15 Sequence 15, Appli
7	216.8	21.4	1110	3	US-09-823-373-16 Sequence 16, Appli
8	132.4	13.1	1194	2	US-08-447-702-4 Sequence 4, Appli
9	132.4	13.1	1194	2	US-08-465-615-4 Sequence 3, Appli
10	84	8.3	385	3	US-09-823-373-3 Sequence 66400, A
11	36.6	3.6	601	3	US-09-949-016-66400 Sequence 13698, A
12	36.6	3.6	110243	3	US-09-949-016-13698 Sequence 30, Appli
13	35.4	3.5	90	3	US-09-823-373-30 Sequence 27, Appli
14	35	3.5	90	3	US-09-823-373-32 Sequence 32, Appli
15	34.8	3.4	90	3	US-09-823-373-32 Sequence 60707, A
16	34.2	3.4	601	3	US-09-949-016-60707 Sequence 546, App
17	34.2	3.4	605	3	US-09-270-767-546 Sequence 15828, A
18	34.2	3.4	605	3	US-09-270-767-546 Sequence 2, Appli
19	34.2	3.4	1434	3	US-09-491-772-2 Sequence 1, Appli
20	34.2	3.4	8580	3	US-09-491-772-1 Sequence 13506, A
21	34.2	3.4	283538	3	US-09-949-016-13506 Sequence 60708, A
22	33.8	3.3	601	3	US-09-949-016-60708 Sequence 543, App
23	33	3.3	1401	4	US-09-605-703B-543 Sequence 3762, App
24	32.8	3.2	1068	3	US-09-949-016-3762

C	25	32.8	3.2	1115	3	US-09-949-016-539	Sequence 539, App
C	26	32.8	3.2	2088	3	US-09-023-655-340	Sequence 340, App
	27	32.6	3.2	336	3	US-09-643-597-286	Sequence 286, App
	28	32.6	3.2	336	3	US-09-480-884A-286	Sequence 286, App
	29	32.6	3.2	336	3	US-09-542-615A-286	Sequence 286, App
	30	32.6	3.2	336	3	US-09-606-421B-286	Sequence 286, App
	31	32.6	3.2	336	3	US-09-630-940B-286	Sequence 286, App
	32	32.6	3.2	336	3	US-10-007-700-286	Sequence 286, App
	33	32.6	3.2	2835	3	US-09-134-001C-1515	Sequence 1515, App
	34	32.6	3.2	3561	3	US-09-134-001C-1685	Sequence 1685, App
C	35	32.6	3.2	6414	3	US-09-134-001C-1626	Sequence 1626, App
	36	32.4	3.2	561	3	US-09-902-540-6607	Sequence 6607, App
	37	32.4	3.2	2260	3	US-09-902-540-502	Sequence 502, App
C	38	32.2	3.2	505	3	US-09-621-976-15639	Sequence 15639, App
C	39	32	3.2	2652	2	US-08-318-831-1	Sequence 1, Appli
	40	31.8	3.1	894	3	US-09-489-039A-5631	Sequence 5631, App
C	41	31.8	3.1	1930	3	US-09-270-767-9205	Sequence 9205, App
C	42	31.8	3.1	1930	3	US-09-270-767-24487	Sequence 24487, A
	43	31.6	3.1	29653	3	US-09-949-016-13442	Sequence 13442, A
	44	31.6	3.1	37304	3	US-09-949-016-11862	Sequence 11862, A
	45	31.6	3.1	481115	3	US-09-949-016-11940	Sequence 11940, A

ALIGNMENTS

RESULT 1  
US-08-690-493-2  
Sequence 2, Application US/08690493  
Patent No. 5872000  
GENERAL INFORMATION:  
APPLICANT: Yu, Fujio  
TITLE OF INVENTION: No. 5872000e1 Nitrilase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steinberg, Raklin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,493  
FILING DATE: 31 JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 213061/1995  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 3821005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 768-3800  
TELEFAX: (212) 382-2124  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1035 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Gordona terrae  
STRAIN: MA-1  
CELL TYPE: unicellular organism  
US-08-690-493-2  
Query Match 27.9%; Score 283; DB 2; Length 1035;







TYPE: DNA  
ORGANISM: Acidovorax facilis  
US-09-823-373-4

Query Match 21.5%; Score 218; DB 3; Length 1110;  
Best Local Similarity 52.0%; Pred. No. 4.6e-61;  
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

```
QY 22 GCCTGCGTGAAGCCGCCCCGATCTACATGATTGGAGCGGACGGTGACAAACCATT 81
Db 31 GCAACCGTTGAGCAGACCGGGTATGGCTCGACGAGACGCAACGATCGACAGTGCATC 90
QY 82 GAGTTGATGAAGAAGCAGCACTAATATGCTGCTGATCGCTTCCGGAACCTTGG 141
Db 91 GGATCATCGAAGAAGCTGCCCAAAAGGCGGAGTCTGATCGCTTCCGGAAGTATTC 150
QY 142 ATTCCAGGCTACCCATGCTTCTTGGCTTGACTCAACGATGGCAATGCAATTGTA 201
Db 151 ATTCCGGGCTACCCCTATTGGGCGGTGGCTCGGCGAGCTGAAGTACAGCCTAAGCTTACT 210
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCATTTAGAT 261
Db 211 TCACGCTATCACGAGAAATCGTTGGAGCTAGGTGACGACCGTATGCGTCCGCTCAGCTG 270
QY 262 GCAGCCCAAGCGGTTGGGATCATGGTCACCCCTGGGGATGATGAACGGGTGGGCAAC 321
Db 271 GCCGCGCGCCGCAACAATCGCACTGTCATGGGCTATTTCGGAAGCGGGAAGCCGATCG 330
QY 322 CTTACATCAGTCAGTGGTTCATAGGCGATTAATGTTGACACCATTTGGGGCCCGGCGAAG 381
Db 331 CGCTATCTGAGCCAGGTGTTTCATCGACGAGCGTGGCGAGATCGTTCCAATCGGCGCAAG 390
QY 382 TTGAAACCTTACTTTTGTGAAGTACTTTGTTCCGCGAAGGGGATGTTTCATCGTACGG 441
Db 391 CTGAAGCCCAACACACGTTGAGGCTACGATCTACGCGCAAGCAACGAAACGATTTCTC 450
QY 442 GTTTTCGAGACGTCGTTGGAAGCGTGGGCTTATGCTGTGGAGCACTTCAACCG 501
Db 451 ACGCAGCACTTCGCGTTCGAGCGCGTGGTGAATGAACGCTGGGAACATTTCCAAACCG 510
QY 502 CTAAACAATAACGCTTTGTATGCACAAATGAAGAGATTCATTGCGGCTTGCGCGAGC 561
Db 511 CTGACGAAGTTTCATGATGTAAGCCTCGGTGAGCAGGTCCACGTTGCATCGTGGCGGCG 570
QY 562 TTAGCCTTTATCCTAATGCGGCGAAGCCCTGGGCGCTGATGTCATGTAGCGGCTCT 621
Db 571 ATGTCCCTCTTTCAGCCGAGATGTTTCCAACTGAGCATCGAAGCCCAACGCGATCAC 630
QY 622 CGAATCTATGCCGTTGAAGGCAATGCTTCTGCTAGCGGTGCTGCGCTCGTTTCAACA 681
Db 631 CGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGCTGACGCAAGGTGATCGGACT 690
QY 682 TCCATGATGATATGCTTGTATGACATGACGAAAGCATGCGTTGCTTCTGGCTGTGCT 741
Db 691 AGCGGATCGAAACGTTCTGCTCAACGACGAAACAGCGCGCACTGTTGCCGCAAGATGT 750
QY 742 GGAAGTCAAGTATCATAGGGGCTGATGGTGTGACTTGTGCGGCTCTTGCGGAAAT 801
Db 751 GGCTGGGCGCGCATTTACGGCCCGGATGAAAGCGAGCTTGCGAAGCCTCTGCGGAGAT 810
QY 802 GAAGAGGTATTTCTCTACGAAACCTTGATCCTGAGTACGCAATCCTTGTAAATGGCG 861
Db 811 GCTGAGGGGATCTTGTACGAGAGATCGATCTGAGCAGATTTGCTGGCGAAGGCTGGA 870
QY 862 GCAGACCTGCTGCTCATTTATCCGTTCCGACATTAATCGCTTGTGCTAATAGATCGAGC 921
Db 871 GCCGATCGGTCGGGCACTATTCGGGCGCTGACGTGCTGCTGCTCAGTTCCAGCCGCGC 930
QY 922 CCTAATTAACCGGTAGTGAATTTGAAGTGAATCTTCG 959
Db 931 AATCATACGCCAGTTTCATCGCATCGGATGACGGTGC 968
```

RESULT 5

US-09-823-373-13  
Sequence 13, Application US/09823373

Patent No. 6870038

GENERAL INFORMATION:

APPLICANT: Chauhan, Sarita

APPLICANT: DiCosimo, Robert

APPLICANT: Payne, Mark

APPLICANT: Gavagan, John

APPLICANT: Fallon, Robert

TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from

TITLE OF INVENTION: Acidovorax Facilis 72W

FILE REFERENCE: BC-1032 US NA

CURRENT FILING DATE: 2001-03-30

PRIOR APPLICATION NUMBER: 60/193,707

PRIOR FILING DATE: 2000-03-31

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Microsoft Office 97

SEQ ID NO 13

LENGTH: 1110

TYPE: DNA

ORGANISM: Acidovorax facilis

US-09-823-373-13

Query Match 21.5%; Score 218; DB 3; Length 1110;  
Best Local Similarity 52.0%; Pred. No. 4.6e-61;  
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

```
QY 22 GCCTGCGTGAAGCCGCCCCGATCTACATGATTGGAGCGGACGGTGACAAACCATT 81
Db 31 GCAACCGTTGAGCAGACCGGGTATGGCTCGACGAGACGCAACGATCGACAGTGCATC 90
QY 82 GAGTTGATGAAGAAGCAGCACTAATATGCTGCTGATCGCTTCCGGAACCTTGG 141
Db 91 GGATCATCGAAGAAGCTGCCCAAAAGGCGGAGTCTGATCGCTTCCGGAAGTATTC 150
QY 142 ATTCCAGGCTACCCATGCTTCTTGGCTTGACTCAACGATGGCAATGCAATTGTA 201
Db 151 ATTCCGGGCTACCCCTATTGGGCGGTGGCTCGGCGAGTGAAGTACAGCCTAAGCTTACT 210
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCATTTAGAT 261
Db 211 TCACGCTATCACGAGAAATCGTTGGAGCTAGGTGACGACCGTATGCGTCCGCTCAGCTG 270
QY 262 GCAGCCCAAGCGGTTGGGAAATCATGCTCACCCCTGGGGATGATGAACGGGTGGGCAAC 321
Db 271 GCCGCGCGCCGCAACAATCGCACTGCTCATGGGCTATTTCGAGCGGGAAGCCGATCG 330
QY 322 CTTACATCAGTCAGTGGTTCATAGGCGATTAATGTTGACACCATTTGGGGCCCGGCGAAG 381
Db 331 CGCTATCTGAGCCAGGTGTTTCATCGACGAGCGTGGCGAGATCGTTCCAATCGGCGCAAG 390
QY 382 TTGAAACCTTACTTTTGTGAAGTACTTTGTTCCGCGAAGGGGATGTTTCATGCTAGCG 441
Db 391 CTGAAGCCCAACACACGTTGAGCGGTACGATCTACGCGAAGGCAACGGAACGATTTCTC 450
QY 442 GTTTTCGAGACGTCGTTGGAAGCGTGGGCTTATGCTGTGGAGCACTTCAACCG 501
Db 451 ACGCAGCACTTCGCGTTCGAGCGCGTGGTGAATGAACGCTGGGAACATTTCCAAACCG 510
QY 502 CTAAACAATAACGCTTTGTATGCACAAATGAAGAGATTCATTGTGCGGCTTGCGCGAGC 561
Db 511 CTGACGAAGTTTCATGATGTAAGCCTCGGTGAGCAGGTCCACGTTGCATCGTGGCGGCG 570
QY 562 TTAGCCTTTATCCTAATGCGGCGAAGCCCTGGGCGCTGATGTCATGTAGCGGCTCT 621
Db 571 ATGTCCCTCTTTCAGCCGAGATGTTTCCAACTGAGCATCGAAGCCCAACGCGATCAC 630
QY 622 CGAATCTATGCCGTTGAAGGCAATGCTTGTGCTAGCGTGTGCGCTCGTTTCAACA 681
Db 631 CGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGCTGACGCAAGGTGATCGGACT 690
```

```
QY      682 TCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTCTGCGTGTGT 741
        |||||
Db      691 AGCGGATCGAAACGTTCTGCTCAACGACGACGCGCATGTTGCCGAAGATGT 750
QY      742 GGAACCTCAGCTATCATAGGCGCTGATGCTGTGATGCTGTGCGCCTCTTGCCGAAAT 801
        |||||
Db      751 GCGTGGGCGCGCATTTACGCGCCGATGGAAGCGAGCTTGGGAAGCCTCTGCGGAAGAT 810
QY      802 GAAGAGGTATTTCTTACGCAAACTTGATCCTGAGTACGCATCCTTGCTAAATGGCG 861
        |||||
Db      811 GCTGAGGGGATCTTGTACGACAGATCGATCTGGAGCAGATTCTGCTGGGAAGCTGGA 870
QY      862 GCAGACCTCTGCTGTGCTATTATCCCGTCCCGACATTACTCGCTTGCTAATAGATCGCAGC 921
        |||||
Db      871 GCGATCCGGTGGGCGCACTATTCGGCGCTGACGTGCTGCTCCAGTTCCAGTTCGACCCGCGC 930
QY      922 CCTAAATTACCGGTAGTTGAAATGAAGGTGATCTTCG 959
Db      931 AATCATACGCCAGTTTCATCGCATTCGCATTGACGCTCG 968
```

## RESULT 6

```
US-09-823-373-15
; Sequence 15, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: Dicosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Acidovorax delafieldii
US-09-823-373-15
```

Query Match 21.5%; Score 218; DB 3; Length 1776;  
Best Local Similarity 52.0%; Pred. No. 6e-61;  
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

```
QY      22 GCCTGCGTGCAAGCGCCCGATCTACATGATTTGAGCGGACGCTGGACAAAACCAT 81
        |||||
Db      362 GCAACGTTACGACGAGCCGGTATGGCTCGACGACGACGACGATCGACAAGTGAATC 421
QY      82 GAGTTGATGGAAGAAGACGACGTAATATGCTCTGTCGTCCTTCCGAAACTGG 141
        |||||
Db      422 GGCATCATCGAAGAAGCTGCCCAAAAGGGCGGAGTCTGATCGCTTCCCGAAGTATTC 481
QY      142 ATTCAGGCTTACCCTATGTTCTTTGGCTTGACTCACCAGCATGGGCAATGCAATTGTA 201
        |||||
Db      482 ATTCCGGGCTACCCCTATTGGGGGTGGCTCGGCGACGTGAAGTACAGCCTAAGCTTACT 541
QY      202 CGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCATTTAGAT 261
        |||||
Db      542 TCACGCTATCACGAGAATTGTTGAGCTAGGTAGCAGACCGTATGCGCTCCAGCTG 601
QY      262 GCAGCCAAAGCGGTTGGAAATCATGTCAACCTGGGAGTGAAGACGGGTGCTGCGACC 321
        |||||
Db      602 GCCGCGCGCGCAAAATCGCACTGTCATGCGCTATTGCGAGCGGGAAGCCGGAATCG 661
QY      322 CTTTACATCAGTCAGTGTTCATAGGCGATAATGTTGACACCATTTGGGCGCGGGAAG 381
        |||||
Db      662 CGCTATCTGAGCGCAGGTGTTCATCGACGAGCGGTGCGAGATGTTGCCAATCGGCGGAG 721
```

```
QY      382 TTGAACCTACTTTTGTGAACGTACTTGTGTCGGCGAAGGGATGTTTCATCGCTAGCG 441
        |||||
Db      722 CTGAAGCCACACACGTTGAGCGGTACGATCTACGGCGAAGGCAACGGAACCGATTCTC 781
QY      442 GTTTTCGAGACGCTCTGTGGAAGGCTGGGTGCTTATGCTGTTGGAGCACCTTCAACCG 501
        |||||
Db      782 ACGACGACTTCGCGCTTCGGAACGCGTGGGTGATGAACTGCTGGGAACATTTCCAAACCG 841
QY      502 CTAACAAATACGCTTTGTATGACAAATGAAGATTCATTTGCGGCTTGCGCGAGC 561
        |||||
Db      842 CTCAGCAAGTTCATGATGTAACAGCTCGGTGAGCAGGTCCACGTTGCATCGTGCCGCG 901
QY      562 TTAGCCTTATCTTAATGCGCGGAAGCCCTGGGCTGATGTCAATGTAAGCGCTCT 621
        |||||
Db      902 ATGTCCTCTTCAGCCGATGTTTCCAACTGAGCATCGAAGCCAACGCGAGCTCAC 961
QY      622 CGAATCTATCCGTTGAAGGCAATGCTTCGTACTAGCGTGTGCGCTCGTTTCAACA 681
        |||||
Db      962 CGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTCTGACGCAAGGTGATCGACCT 1021
```

```
QY      682 TCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTCTGCGTGTGT 741
        |||||
Db      1022 AGCGGATCGAAACGTTCTGCTCAACGACGACGCGCACTGTTGCCGAAGATGT 1081
QY      742 GGAACCTCAGCTATCATAGGCGCTGATGCTGTGCTGTGCGCCTCTTGCCGAAAT 801
        |||||
Db      1082 GCGTGGGCGCGCATTTACGCGCCGATGGAAGCGAGCTTGCAAGCCTTGCGCGAAGAT 1141
QY      802 GAAGAGGTATTTCTTACGCAAACTTGATCCTGAGTACGCATCCTTGCTAAATGGCG 861
        |||||
Db      1142 GCTGAGGGGATCTTGTACGACAGATCGATCTGAGCAGATTCTGCTGCGAAGCTGGA 1201
QY      862 GCAGACCTGCTGTGCTATTTCCCGTCCCGACATTACTCGCTTGCTAATAGATCGCAGC 921
        |||||
Db      1202 GCGGATCCGGTGGGCACTATTGCGGCGCTGACGTGCTGCTCAGTTGACCCGCGC 1261
QY      922 CCTAAATTACCGGTAGTTGAAATGAAGGTGATCTTCG 959
Db      1262 AATCATACGCCAGTTTCATCGCATTCGCATTGACGCTCG 1299
```

## RESULT 7

```
US-09-823-373-16
; Sequence 16, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: Dicosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: A synthetic
; OTHER INFORMATION: version of the nitrilase gene
US-09-823-373-16
```

Query Match 21.4%; Score 216.8; DB 3; Length 1110;  
Best Local Similarity 51.9%; Pred. No. 1.1e-60;  
Matches 488; Conservative 0; Mismatches 452; Indels 0; Gaps 0;





```
QY 314 GTGGACCCCTTACATCAGTCAGTGTTCATAGCGCATATGTGACACCATTTGGGCCC 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 ATGCTCGCTATATTGACGCAATTTGGTTGACCCGAAATGTAATTGATTGGCAAGC 465
QY 374 GGGGAAAGTTGAAACCTTACTTTTGTGAAACGTAATTGTTGGCGAAGGGAGTTTCAT 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 ACAGGAAATTCAGAGCCCACTAGTAGTAAGAAAGAGCTGTATGGGAGATGGGAGTGAAGCA 525
QY 434 CGCTAGCGGTTTTCGACGCTGTGTGGAAGCGTGGGCTTATGCTGTGGGAGCACC 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 TGGCTCCCGTATTAAACAGAGTATGGGAATCTTGGGGAGCTCCAGTGTGGGACATG 585
QY 494 TTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGATTCATTGTGGGCTT 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 CTCTCCCATTAACATTCGCGCGATGGGCTCATTTGAACGAACAGTACATGTGCTTCT 645
QY 554 GGGCGAGCTTTAGCCTTATCTTAATGCGCGGAAGCCCTGGGCGTATGTCATGTAG 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 GGGCAGCCTTCGTCCTTAAGCGCAGTATCATCCAGAGTATCATCCAGCCTGTGCGGT 705
QY 614 GGGCCTCTCGAATCTATGCGGCTGAAGGCAATGCTC-----G 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 CTACTAATGCGATGATCAGATCATTAGTCAAGTTTAACGCGATCAGCAATCAGGTATATG 765
QY 653 TACTAGCGTGTGCGCTGTTTTCACAATCCATGATGATGCTTTGTACAGATGACG 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 TAATATGTCAACCAATCTCGTTGGCCAAGACATGATGATGATGGGAAGATGAAT 825
QY 713 AAAAGCATGCGTTCCTTGCTGTGTGTGACATCAAGTATCATAGGCGCTGATGTG 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 TTTCAAAACCTTTCTACCGCTTGTTCTGGAACACAGCGATTAT---TTCTAACACCG 882
QY 773 GTGACTGTGCGCGCTCTTGCCGAAATGAAGAGGTAATCTCTACGCAACCTTGATC 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GTGAGATTTGGCATCAATTCACAAAGACGCGAGGGAATGCTGTGACAGATTGACC 942
QY 833 CTGAGTAGCATCTTGTCTAAATGCGGCGAGACCTGCTGTCATTATTCCTGCCG 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 TTAACCAATTAATTATGAAAGTGTTACTGTGATCCCGCGGTCACTTACTTCCCG 1002
QY 893 ACATTACTGCTGTCTAATAGATCGACGCCCTAAATTAACCGTAGTTGAATG 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 GCTTCTTAAGTTGACATTTGATCAGTCTGAACATGTAACCGTAACCAATAG 1056
```

RESULT 9  
US-08-465-615-4  
; Sequence 4, Application US/08465615  
; Patent No. 5635391

; GENERAL INFORMATION:  
; APPLICANT: PETRE, Dominique  
; APPLICANT: CERBELEAUD, Edith  
; APPLICANT: LEVY-SCHIL, Sophie  
; APPLICANT: CROUZET, Joel  
; TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
; TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
; TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
; TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
; TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Burns, Doane, Swecker & Mathis  
; STREET: P.O. Box 1404  
; CITY: Alexandria  
; STATE: Virginia  
; COUNTRY: United States  
; ZIP: 22313-1404  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/465,615  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/194,588  
; FILING DATE: 10-FEB-1994  
; APPLICATION NUMBER: FR 9209882  
; FILING DATE: 10-AUG-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Crane-Feury, Sharon E  
; REGISTRATION NUMBER: 36,113  
; REFERENCE/DOCKET NUMBER: 003025-015  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-6620  
; TELEFAX: (703) 836-2021  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1194 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-465-615-4
```

Query Match 13.1%; Score 132.4; DB 2; Length 1194;  
Best Local Similarity 48.1%; Pred. No. 7.6e-33;  
Matches 459; Conservative 0; Mismatches 471; Indels 24; Gaps 2;

```
QY 14 TCAAGTCCGCTGCTGCAAGCCGCCCGATCTACATGATTTGGAGCGACGTTGACA 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 106 TCAAGTAGCAGCAGTGCAGAGCTGCTCCGTATTTATGATCTAGAGGCAACAGTAGATA 165
QY 74 AAACATTTGATGATGGAAGAAGCAGCAGTAATATATGCTGCTGATGCTTTCCGG 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 166 AAACCTGTAATTAATAGCAGAAGCAGCATTAATGGGCGCAAGGTTATCGCTTCCAG 225
QY 134 AAACCTTGATTCAGGCTACCCATGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGC 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 226 AAGCATTTATCCCGCTATCCATATTTGGAATTTGACATCAATATGACTTCACTGGA 285
QY 194 AATTTGACGCAATACATGAACTCATTTGATGAGTGGCCCTCAAGCTAAGCGCA 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 286 TGATGAGGCGGCTTTTCAAGAATGCGATGAATCCCAAGCAAGAGTTCAACAAA 345
QY 254 TTTGAGATGAGCCAAAGCGTTGGGAATCATGTCCACCTGGGAGATGAACGGTGC 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 346 TTAGTGATGCTGCAAAAAGAAATGAGTTTACGTTTTCGTTCTGATCAGAGAAAGATA 405
QY 314 GTGGACCCCTTACATCAGTCAGTGTTCATAGCGCATATGTTGACACCATTTGGGCCC 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 406 ATGCTCGCTATATTGACGCAATTTGCTTTGACCCGAAATGTAATTGATTGGCAAGC 465
QY 374 GGGGAAAGTTGAAACCTTACTTTTGTGAAGTACTTTGTTGGCGGAAGGAGTTTCAT 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 466 ACAGGAAATTTAAGCCCACTAGTAGTAAGAGCTGTATGGGAGATGGGAGTGAAGCA 525
QY 434 CGCTAGCGGTTTTCGACGCTGTGTGGAAGCGTGGGCTTATGCTGTGGGAGCACC 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 TGGCTCCCGTATTAAACAGAGTATGGGAATCTTGGGGAGCTCCAGTGTGGGAACATG 585
QY 494 TTCAACCGCTAACAAATACGCTTTGTATGCAAAATGAAGATTCATTGTGGGCTT 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 CTCTCCCATTAACATTCGCGCGATGGGCTCATTTGAACGAACAGTACATGTGCTTCT 645
QY 554 GGGCGAGCTTTAGCCTTATCTTAATGCGCGGAAGCCCTGGGCGTATGTCATGTAG 613
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 GGGCAGCCTTCGTCCTTAAGCGCAGTATCATCCAGAGTATCATCCAGCCTGTGCGGT 705
QY 614 GGGCCTCTCGAATCTATGCGGTTGAAGGCAATGCTC-----G 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 CTACTAATGCGATGATCAGATCATTAGTCAAGTTTACGCGATCAGCAATCAGGTATATG 765
```



```
QY 653 TACTAGCGTGTGCGCTCGTTTCACAATCCATGATCATGCTTTGTACAGATGACG 712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 TAATATGTCAACCAATCTCGTTGGCCAAGACATGATGACATGATGGGAAGATGAAT 825
QY 713 AAAAGCATGCGTGTCTTGTGCTGTGTGACACTCACGTATCATAGGGCTGATGCTG 772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 826 TTTCAAAACCTTCTACCCGCTGTGTTCTGGAACACAGCGATTAT--TTCTAACACCG 882
QY 773 GTGACTTGTGCGCGCTCTTGCCGAATGAAGGGTATTCTCTACGCAACCTTGATC 832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 GTGAGATTGGCATCAATTCCACAAGACCGCGAGGAATTGCTGTTGCAGAGATTGACC 942
QY 833 CTGAGATACGCATCTTGTCTAAATGGCGGACACCTGTGTCATTATTCCTCCCG 892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 TTAACCAATATTTATGGAAGGTGTACTGATCCCGCGGTCAATTACTCTACTCCCG 1002
QY 893 ACATTACTGCTTGTCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAATTG 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 GCTTCTTAAGTTGACATTTGATCAGTCTGAACATGTACCCGTAACCAATAG 1056
```

## RESULT 10

```
US-09-823-373-3
; Sequence 3, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrlase from
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 3
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Acidovorax facilis
US-09-823-373-3
```

```
Query Match      8.3%; Score 84; DB 3; Length 385;
Best Local Similarity 53.3%; Pred. No. 3.6e-17;
Matches 177; Conservative 0; Mismatches 155; Indels 0; Gaps 0;
```

```
QY 208 TACCATGAGACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGATTTCAGATGACGCC 267
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 53 TATCACGAGAATTCGTTGAGCTAGGTGACGACCGTATGCGTCGCCCTCCAGCTGGCCG 112
QY 268 AAGCGTTGGGAATCATGTGTCACCCCTGGGGATGAGTGAACGGGTGGGCAACCTTTAC 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 CGCCGCAACAATCGCACTGTCATGGCTATTGGAAGCGGGAAGCCGGATCGGCTAT 172
QY 328 ATCAGTCAGTGGTTCATAGGCGATTAATGTGACACCATTTGGGCGCGCAAGTTGAAA 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 CTGAGCCAGGTGTTCAATCGACGAGCGTGGCGAGATCGTTGCCAATCGCGCAAGCTGAAG 232
QY 388 CCTACTTTTGTGAAGTACTTTGTTGCGGGAAGGGATGTTTCATCGCTAGCGGTTTTC 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 CCCACACACGTTGAGCGTACGATCTACGCGGAAGCAACGGAACGATTTCCTCACGAC 292
QY 448 GAGACGCTGTGTGAAGCTGGGTGCTTATGCTGTTGGAGACACCTTCAACCGCTAACA 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 GACTTCGCGTTCGAGCGCTCGGTGATTTGAATGCTGCGGAACATTTCCAACGCTCAGC 352
QY 508 AAATACGCTTGTATGACAAATGAAGAGAT 539
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 AAGTTCATGATGTACAGCCTCGGTGAGCAGGT 384
```

## RESULT 11

```
US-09-949-016-66400
; Sequence 66400, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66400
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-66400
```

```
Query Match      3.6%; Score 36.6; DB 3; Length 601;
Best Local Similarity 57.4%; Pred. No. 0.19;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
```

```
QY 259 GATGAGCCAGCGGTTGGAATCATGTCACCCCTGGGATGATGAACGGGTGGTGC 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 GATGAGGAGCGGCTTTGGAGTTGCAACATTTGGTGATGATGAAGTCTTGGGAGC 186
QY 319 ACCCTTACATCAGTCAGTGTTCATAGGCGATATGATGACACCATTTGGGCCC 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 AGTGTACTTAGACAGCAGTGTCTCTCAAGTGTGATCGTGACCCCTGGGGGCTC 241
```

## RESULT 12

```
US-09-949-016-13698
; Sequence 13698, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13698
; LENGTH: 110243
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(110243)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13698
```

```
Query Match      3.6%; Score 36.6; DB 3; Length 110243;
Best Local Similarity 57.4%; Pred. No. 3.3;
Matches 66; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:12:44 ; Search time 976.006 Seconds  
(without alignments)  
8591.290 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaagaagctatcaaggt.....cgcaactcgagaatttga 1014

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA Main:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	1014	100.0	1014	3	US-09-751-299-3 Sequence 3, Appl
2	1012.4	99.8	1014	6	US-10-146-772-385 Sequence 385, App
3	1012.4	99.8	1014	6	US-10-241-742-385 Sequence 385, App
4	1012.4	99.8	1014	7	US-10-440-523-385 Sequence 385, App
5	1012.4	99.8	1014	7	US-10-440-503-385 Sequence 385, App
6	1012.4	99.8	1014	7	US-10-461-925-385 Sequence 385, App
7	499.4	49.3	1014	6	US-10-146-772-47 Sequence 47, Appl
8	499.4	49.3	1014	6	US-10-241-742-47 Sequence 47, Appl
9	499.4	49.3	1014	7	US-10-440-523-47 Sequence 47, Appl
10	499.4	49.3	1014	7	US-10-440-503-47 Sequence 47, Appl
11	499.4	49.3	1014	7	US-10-461-925-47 Sequence 47, Appl
12	497.8	49.1	1014	6	US-10-146-772-57 Sequence 57, Appl
13	497.8	49.1	1014	6	US-10-241-742-57 Sequence 57, Appl
14	497.8	49.1	1014	7	US-10-440-523-57 Sequence 57, Appl
15	497.8	49.1	1014	7	US-10-440-503-57 Sequence 57, Appl
16	497.8	49.1	1014	7	US-10-461-925-57 Sequence 57, Appl
17	399.6	39.4	999	6	US-10-146-772-139 Sequence 139, App
18	399.6	39.4	999	6	US-10-241-742-139 Sequence 139, App
19	399.6	39.4	999	7	US-10-440-523-139 Sequence 139, App
20	399.6	39.4	999	7	US-10-440-503-139 Sequence 139, App
21	399.6	39.4	999	7	US-10-461-925-139 Sequence 139, App
22	396.4	39.1	1002	6	US-10-146-772-291 Sequence 291, App
23	396.4	39.1	1002	6	US-10-241-742-291 Sequence 291, App

24	396.4	39.1	1002	7	US-10-440-523-291	Sequence 291, App
25	396.4	39.1	1002	7	US-10-440-503-291	Sequence 291, App
26	396.4	39.1	1002	7	US-10-461-925-291	Sequence 291, App
27	364.2	35.9	1038	6	US-10-146-772-333	Sequence 333, App
28	364.2	35.9	1038	6	US-10-241-742-333	Sequence 333, App
29	364.2	35.9	1038	7	US-10-440-523-333	Sequence 333, App
30	364.2	35.9	1038	7	US-10-440-503-333	Sequence 333, App
31	364.2	35.9	1038	7	US-10-461-925-333	Sequence 333, App
32	358.6	35.4	1068	6	US-10-146-772-305	Sequence 305, App
33	358.6	35.4	1068	6	US-10-241-742-305	Sequence 305, App
34	358.6	35.4	1068	7	US-10-440-523-305	Sequence 305, App
35	358.6	35.4	1068	7	US-10-440-503-305	Sequence 305, App
36	358.6	35.4	1068	7	US-10-461-925-305	Sequence 305, App
37	319	31.5	1011	6	US-10-146-772-217	Sequence 217, App
38	319	31.5	1011	6	US-10-241-742-217	Sequence 217, App
39	319	31.5	1011	7	US-10-440-523-217	Sequence 217, App
40	319	31.5	1011	7	US-10-440-503-217	Sequence 217, App
41	319	31.5	1011	7	US-10-461-925-217	Sequence 217, App
42	311.4	30.7	1026	6	US-10-146-772-33	Sequence 33, Appl
43	311.4	30.7	1026	6	US-10-241-742-33	Sequence 33, Appl
44	311.4	30.7	1026	7	US-10-440-523-33	Sequence 33, Appl
45	311.4	30.7	1026	7	US-10-440-503-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1  
US-09-751-299-3  
; Sequence 3, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; APPLICANT: Chaplin, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; CURRENT FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
; OTHER INFORMATION: environmental sample  
; NAME/KEY: CDS  
; LOCATION: (1)..(1014)  
US-09-751-299-3

Query Match 100.0%; Score 1014; DB 3; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1014; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAGAAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCTACATGATTGGAG	60
Db	1	ATGAAGAAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCTACATGATTGGAG	60
QY	61	GCGACGCTGACAAACCATTTGATTGATGAAGAAGACGACGTAATATGCTGCTG	120
Db	61	GCGACGCTGACAAACCATTTGATTGATGAAGAAGACGACGTAATATGCTGCTG	120
QY	121	ATGCGCTTCCGAAACTTGGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCA	180
Db	121	ATGCGCTTCCGAAACTTGGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACCA	180



QY 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
Db 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
QY 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
Db 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
QY 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
Db 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
QY 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
Db 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
QY 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
Db 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
QY 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
Db 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTTGCTTCTGCTGTGTGTGACACTCACGTATCATAGGCTGATGTTGACTTG 780  
Db 721 GCGTTGCTTCTGCTGTGTGTGACACTCACGTATCATAGGCTGATGTTGACTTG 780  
QY 781 GTGCGGCTCTTTCGCGAAATGAGAGGGTATTCTTACGCAAACTTGATCTGAGTA 840  
Db 781 GTGCGGCTCTTTCGCGAAATGAGAGGGTATTCTTACGCAAACTTGATCTGAGTA 840  
QY 841 CGCATCTTGTCTAAATGCGGAGACCCCTGCTGTGATTAATCCCGTCCGACATTA 900  
Db 841 CGCATCTTGTCTAAATGCGGAGACCCCTGCTGTGATTAATCCCGTCCGACATTA 900  
QY 901 CGCTTGTCTAATAGATGCGAGCCCTTAATTAACGGTAGTTGAATGAAGTGATCTT 960  
Db 901 CGCTTGTCTAATAGATGCGAGCCCTTAATTAACGGTAGTTGAATGAAGTGATCTT 960  
QY 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA 1014  
Db 961 CCTTACGCTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA 1014

RESULT 2  
US-10-146-772-385  
; Sequence 385, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772

QY 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
Db 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
QY 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
Db 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
QY 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
Db 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
QY 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
Db 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
QY 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
Db 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
QY 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
Db 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720

Query Match 99.8%; Score 1012.4; DB 6; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAGAAGCTATCAAGTTCGCTGCTGCAAGCCGCCGATCTACATGATTGGAG 60  
Db 1 ATGAAGAAGCTATCAAGTTCGCTGCTGCAAGCCGCCGATCTACATGATTGGAG 60  
QY 61 GCGACGCTGACAAACCATTTGATGATGAGAAAGACAGACGTAATATGCTGCTG 120  
Db 61 GCGACGCTGACAAACCATTTGATGATGAGAAAGACAGACGTAATATGCTGCTG 120  
QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTAACCCATGTTTCTTGCTTGACTCACCA 180  
Db 121 ATGCGCTTCCGGAACCTTGATTCAGGCTAACCCATGTTTCTTGCTTGACTCACCA 180  
QY 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
Db 181 GCATGGCAATGCAATTGTACGCCAATACAGAGAACTCATTTGAGTTGGATGGCCCT 240  
QY 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
Db 241 CAAAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGGAATCATGTGACCCCTGGGATG 300  
QY 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
Db 301 AGTGAACGGGTCGGTGGCAACCCCTTACATCAGTCACTGCTTCCATAGGCGAATAGTGAC 360  
QY 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGCAAAAGTTGAACCTACTTTTGTGAACGTAAGTTGTTGGCGAA 420  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTTGTTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
Db 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTGTATGCACAAATAGAGAGATT 540  
QY 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
Db 541 CATTTGCGGCTTGCGCGAGCTTAAAGCTTTATCCTAATGCGGCAAGCCCTGGGGCT 600  
QY 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
Db 601 GATGTCATGTAGCGGCTTCGAATCTATGCGTTGAAGGCAATGCTTCTGTAACG 660  
QY 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTGTTTCAACATCCATGATGATATGCTTGTACAGATGACGAAAGCAT 720

Db 661 TCGTGTGCGCTCGTTTCAACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCAT 720  
QY 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACTATCATAGGGCTGATGTGTGACTTG 780  
Db 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACTATCATAGGGCTGATGTGTGACTTG 780  
QY 781 GTGCGGCTTCTTGCCGAAAATGAAGAGGTATTTCTTACGCAAACTTGATCTGAGTA 840  
Db 781 GTGCGGCTTCTTGCCGAAAATGAAGAGGTATTTCTTACGCAAACTTGATCTGAGTA 840  
QY 841 CGCATCTTGTCTAAATAGCGCGCAGACCTCTGTGTCATTTCCCGTCCGACATTACT 900  
Db 841 CGCATCTTGTCTAAATAGCGCGCAGACCTCTGTGTCATTTCCCGTCCGACATTACT 900  
QY 901 CGCTTGCTAATAGATCGACGCCCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCTG 960  
Db 901 CGCTTGCTAATAGATCGACGCCCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCTG 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014  
Db 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014

RESULT 3  
US-10-241-742-385  
; Sequence 385, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; PRIOR FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 385  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-385

Query Match 99.8%; Score 1012.4; DB 6; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGATCTACATGATTTGAG 60  
Db 1 ATGAAAGAAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGATCTACATGATTTGAG 60  
QY 61 GCGACGCTGACAAACCACTGAGTTGATGGAAGAAGACGACGTAATATGCTGCTG 120

Db 61 GCGACGCTGACAAACCACTGAGTTGATGGAAGAAGACGACGTAATATGCTGCTG 120  
QY 121 ATGCGCTTTCGGAAACTTGGATTCCAGGCTACCCATGGTTTCTTTGGCTTGACTCACCA 180  
Db 121 ATGCGCTTTCGGAAACTTGGATTCCAGGCTACCCATGGTTTCTTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAGATGACAGCCAGCGTTGGAAATCATGTGTCAACCTGGGAGT 300  
Db 241 CAAGCTAAGCGCATTTCAGATGACAGCCAGCGTTGGAAATCATGTGTCAACCTGGGAGT 300  
QY 301 AGTGAACGGGTGCGGTGCAACCTTTACATCAGTCAGTGGTTTCATAGCGGATAATGTGAC 360  
Db 301 AGTGAACGGGTGCGGTGCAACCTTTACATCAGTCAGTGGTTTCATAGCGGATAATGTGAC 360  
QY 361 ACCATTGGGGCCCCGGCGAAAGTTGAAACCTACTTTTGTGAAACGTACTTTGTTGGCGAA 420  
Db 361 ACCATTGGGGCCCCGGCGAAAGTTGAAACCTACTTTTGTGAAACGTACTTTGTTGGCGAA 420  
QY 421 GGGATGCTTCATCGCTACGCGGTTTTCGAGACGTCGTGTGAAAGGCTGGGTGCTTATGC 480  
Db 421 GGGATGCTTCATCGCTACGCGGTTTTCGAGACGTCGTGTGAAAGGCTGGGTGCTTATGC 480  
QY 481 TGTGGAGACACTTCAACCGCTAACAATAACGCTTGTATGACAAATGAAGATTT 540  
Db 481 TGTGGAGACACTTCAACCGCTAACAATAACGCTTGTATGACAAATGAAGATTT 540  
QY 541 CATTTGCGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCT 600  
Db 541 CATTTGCGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAAGGCAATGCTTCTGACTAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAAGGCAATGCTTCTGACTAGCG 660  
QY 661 TCGTGTGCGCTGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCAT 720  
Db 661 TCGTGTGCGCTGTTTCACAATCCATGATGATATGCTTTGTACAGATGACGAAAAGCAT 720  
QY 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACTACGTAATAGGGCTGATGTGTGACTTG 780  
Db 721 GCGTTGCTTCTGGCTGTGTGTGACACTCACTACGTAATAGGGCTGATGTGTGACTTG 780  
QY 781 GTGCGGCTTCTTGCCGAAAATGAAGAGGTATTTCTTACGCAAACTTGATCTGAGTA 840  
Db 781 GTGCGGCTTCTTGCCGAAAATGAAGAGGTATTTCTTACGCAAACTTGATCTGAGTA 840  
QY 841 CGCATCTTGTCTAAATAGCGCGCAGACCTCTGTGTCATTTCCCGTCCGACATTACT 900  
Db 841 CGCATCTTGTCTAAATAGCGCGCAGACCTCTGTGTCATTTCCCGTCCGACATTACT 900  
QY 901 CGCTTGCTAATAGATCGACGCCCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCTG 960  
Db 901 CGCTTGCTAATAGATCGACGCCCCCTAAATTACCGGTAGTTGAAATTGAAGGTGATCTTCTG 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014  
Db 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014

RESULT 4  
US-10-440-523-385  
; Sequence 385, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen

APPLICANT: Milan, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Madden, Mark  
APPLICANT: Burk, Mark  
TITLE OF INVENTION: Nitrlases  
FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
CURRENT APPLICATION NUMBER: US/10/440,523  
CURRENT FILING DATE: 2003-05-15  
PRIOR APPLICATION NUMBER: US/10/146,772  
PRIOR FILING DATE: 2002-05-15  
PRIOR APPLICATION NUMBER: US 60/309,006  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US 60/351,336  
PRIOR FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: US 60/300,189  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 09/751,299  
PRIOR FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 385  
LENGTH: 1014  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-385

Query Match 99.8%; Score 1012.4; DB 7; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAAGCTATCAAGTCGCCCTGCGTGCAAGCCCGCCGATCTACATGGATTGGAG 60  
DB 1 ATGAAAGAAGCTATCAAGTCGCCCTGCGTGCAAGCCCGCCGATCTACATGGATTGAA 60  
QY 61 GCGACGGTGACAAAACCATTTGATGATGGAAGAACGACGCACTAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAAACCATTTGATGATGGAAGAACGACGCACTAATATGCTGCTG 120  
QY 121 ATGCGCTTTCGGAACCTTGATTCAGGCTACCCAGTGTCTTTGCTTGACTCACCA 180  
DB 121 ATGCGCTTTCGGAACCTTGATTCAGGCTACCCAGTGTCTTTGCTTGACTCACCA 180  
QY 181 GCATGGCAATGCAATTTGATGCGCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
DB 181 GCATGGCAATGCAATTTGATGCGCAATACCATGAGAACTCATTTGAGTTGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAAGTCAGCCAGCGGTTGGGAATCATGTGTCACCCCTGGGATG 300  
DB 241 CAAGCTAAGCGCATTTCAAGTCAGCCAGCGGTTGGGAATCATGTGTCACCCCTGGGATG 300  
QY 301 AGTGAACGGGTGCGTGGCAACCTTTACATCAGTCAGTGGTTTATAGGCGAATATGTGAC 360  
DB 301 AGTGAACGGGTGCGTGGCAACCTTTACATCAGTCAGTGGTTTATAGGCGAATATGTGAC 360  
QY 361 ACCATTGGGGCGCGGGAAGTTGAAACCTACTTTTGTGAAACGTACTTTGTTGGGGAA 420  
DB 361 ACCATTGGGGCGCGGGAAGTTGAAACCTACTTTTGTGAAACGTACTTTGTTGGGGAA 420  
QY 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGTTGGAAGGCTGGGCTTATGC 480  
DB 421 GGGGATGTTTCATCGCTAGCGGTTTTCGAGACGTCGTGTTGGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGACCTTCAACCGCTTAACAAATATAGCTTTGTATGCAAAATGAAGAGATT 540  
DB 481 TGTGGGAGACCTTCAACCGCTTAACAAATATAGCTTTGTATGCAAAATGAAGAGATT 540  
QY 541 CATTTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGGCGGGAAGCCCTGGGCT 600

DB 541 CATTTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGGCGGGAAGCCCTGGGCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
DB 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTACTAGCG 660  
QY 661 TCGTGTGCGCTGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
DB 661 TCGTGTGCGCTGTTTCACAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTTGCTTGGCTGTGTGTGACACTCAGTATCATAGGGGCTGATGGTGTGACTTG 780  
DB 721 GCGTTGCTTGGCTGTGTGTGACACTCAGTATCATAGGGGCTGATGGTGTGACTTG 780  
QY 781 GTCGCGCTCTTGCCGAAATGAAGGGTATTCTTACGCAAACTTGATCCTGAGTA 840  
DB 781 GTCGCGCTCTTGCCGAAATGAAGGGTATTCTTACGCAAACTTGATCCTGAGTA 840  
QY 841 CGCATCCTTGCTAAATGCGGACAGCCCTGCTGTCATTATCCGTCGACATTACT 900  
DB 841 CGCATCCTTGCTAAATGCGGACAGCCCTGCTGTCATTATCCGTCGACATTACT 900  
QY 901 CGCTGTATATAGATCGACGCCCTAATTAACCGGTAGTTGAAATGAAGGTGATCTTCT 960  
DB 901 CGCTGTATATAGATCGACGCCCTAATTAACCGGTAGTTGAAATGAAGGTGATCTTCT 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA 1014  
DB 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA 1014

RESULT 5  
US-10-440-503-385  
Sequence 385, Application US/10440503  
Publication No. US20040038419A1  
GENERAL INFORMATION:  
APPLICANT: Weiner, David Paul  
APPLICANT: Chaplin, Jennifer Ann  
APPLICANT: Chi, Ellen  
APPLICANT: Milan, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Burk, Mark J.  
APPLICANT: McQuaid, Jeffrey  
APPLICANT: Stege, Justin  
TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
TITLE OF INVENTION: NITRILES AND/OR CYANIDE  
FILE REFERENCE: 09010-900001  
CURRENT APPLICATION NUMBER: US/10/440,503  
CURRENT FILING DATE: 2003-05-15  
PRIOR APPLICATION NUMBER: US 60/380,737  
PRIOR FILING DATE: 2002-05-15  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 385  
LENGTH: 1014  
TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-385

Query Match 99.8%; Score 1012.4; DB 7; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAAGCTATCAAGTCGCCCTGCGTGCAAGCCCGCCGATCTACATGGATTGGAG 60  
DB 1 ATGAAAGAAGCTATCAAGTCGCCCTGCGTGCAAGCCCGCCGATCTACATGGATTGAA 60  
QY 61 GCGACGGTGACAAAACCATTTGATGATGGAAGAACGACGCACTAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAAACCATTTGATGATGGAAGAACGACGCACTAATATGCTGCTG 120



QY 121 ATGCGCTTTCGGAAACTTGGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
Db 121 ATCGCCTTTCGGAAACTTGGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATGTGAGTTGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATGTGAGTTGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAAGATGCAAGCCGAGGTTGGAAATCATGGTCACCCCTGGGAGT 300  
Db 241 CAAGCTAAGCGCATTTCAAGATGCAAGCCGAGGTTGGAAATCATGGTCACCCCTGGGAGT 300  
QY 301 AGTGAACGGGTCCGTGGCAACCCCTTTACATCAGTCAAGTGTTCATAGGCCATATGTGTAC 360  
Db 301 AGTGAACGGGTCCGTGGCAACCCCTTTACATCAGTCAAGTGTTCATAGGCCATATGTGTAC 360  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTGAAGTACTTTGTTCCGGCAA 420  
Db 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTGAAGTACTTTGTTCCGGCAA 420  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGCACCTTCAACCCGTTAACAATAACGCTTTGTATGACAAATGAAGAGATT 540  
Db 481 TGTGGGAGCACCTTCAACCCGTTAACAATAACGCTTTGTATGACAAATGAAGAGATT 540  
QY 541 CATGTGCGGCTTGCGCGAGCTTACCTTTATCCTAATGCGCGAAAGCCCTGGGCT 600  
Db 541 CATGTGCGGCTTGCGCGAGCTTACCTTTATCCTAATGCGCGAAAGCCCTGGGCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGATAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGATAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCATGATGATGCTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCATGATGATGCTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTGTCTTCTGCGTGTGTGACACTCAGTATCATAGGCGCTGATGCTGATCTG 780  
Db 721 GCGTGTCTTCTGCGTGTGTGACACTCAGTATCATAGGCGCTGATGCTGATCTG 780  
QY 781 GTCCGCGCTTGTGCCGAAATGAGAAGGATTTCTCTACGCAACCTTGATCCTGAGTA 840  
Db 781 GTCCGCGCTTGTGCCGAAATGAGAAGGATTTCTCTACGCAACCTTGATCCTGAGTA 840  
QY 841 CGCATCTTGTCTAAATGCGGAGACCTGCTGTGTCATTAATCCCGTCCGACATTAAT 900  
Db 841 CGCATCTTGTCTAAATGCGGAGACCTGCTGTGTCATTAATCCCGTCCGACATTAAT 900  
QY 901 CGCTTGCTAATAGATCGCAGCCCTTAATTAACCGTAGTTGAATGAAGGTGATCTTCT 960  
Db 901 CGCTTGCTAATAGATCGCAGCCCTTAATTAACCGTAGTTGAATGAAGGTGATCTTCT 960  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAATTGTA 1014  
Db 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAATTGTA 1014

RESULT 6  
US-10-461-925-385  
; Sequence 385, Application US/10461925  
; Publication No. US20040053378A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001

; CURRENT APPLICATION NUMBER: US/10/461,925  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,317  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/392,944  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 385  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-385

Query Match 99.8%; Score 1012.4; DB 7; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAGAAGCTATCAAGTTCGCTCGGTGCAAGCCGCCGATCTACATGATTTGAG 60  
Db 1 ATGAAGAAGCTATCAAGTTCGCTCGGTGCAAGCCGCCGATCTACATGATTTGAG 60  
QY 61 GCGACGGTGGACAAACCAATTGAGTGTGATGGAAGACGACGATTAATGCTGCTG 120  
Db 61 GCGACGGTGGACAAACCAATTGAGTGTGATGGAAGACGACGATTAATGCTGCTG 120  
QY 121 ATGCGCTTTCGGAAACTTGGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
Db 121 ATGCGCTTTCGGAAACTTGGATTCAGGCTACCCATGGTTCTTTGGCTTGACTCACCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATGTGAGTTGATGGCCCT 240  
Db 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATGTGAGTTGATGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAAGATGCAAGCCGAGGTTGGAAATCATGGTCACCCCTGGGAGT 300  
Db 241 CAAGCTAAGCGCATTTCAAGATGCAAGCCGAGGTTGGAAATCATGGTCACCCCTGGGAGT 300  
QY 301 AGTGAACGGGTCCGTGGCAACCCCTTTACATCAGTCAAGTGTTCATAGGCCATATGTGTAC 360  
Db 301 AGTGAACGGGTCCGTGGCAACCCCTTTACATCAGTCAAGTGTTCATAGGCCATATGTGTAC 360  
QY 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTGAAGTACTTTGTTCCGGCAA 420  
Db 361 ACCATTGGGGCCCGCGAAAGTTGAACCTACTTTGTGTGAAGTACTTTGTTCCGGCAA 420  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGAAGGCTGGGCTTATGC 480  
Db 421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGAAGGCTGGGCTTATGC 480  
QY 481 TGTGGGAGCACCTTCAACCGCTTAACAATAACGCTTGTATGCACAAAATGAAGAGATT 540  
Db 481 TGTGGGAGCACCTTCAACCGCTTAACAATAACGCTTGTATGCACAAAATGAAGAGATT 540  
QY 541 CATTTGCGGCTTGCGCGAGCTTACCTTTATCTTAATGCGCGAAAGCCCTGGGCT 600  
Db 541 CATTTGCGGCTTGCGCGAGCTTACCTTTATCTTAATGCGCGAAAGCCCTGGGCT 600  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGATAGCG 660  
Db 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCAATGCTTCTGATAGCG 660  
QY 661 TCGTGTGCGCTCGTTTCACAATCATGATGATGCTTGTACAGATGACGAAAGCAT 720  
Db 661 TCGTGTGCGCTCGTTTCACAATCATGATGATGCTTGTACAGATGACGAAAGCAT 720  
QY 721 GCGTGTCTTCTGCGTGTGTGACACTCAGTATCATAGGCGCTGATGCTGATCTG 780  
Db 721 GCGTGTCTTCTGCGTGTGTGACACTCAGTATCATAGGCGCTGATGCTGATCTG 780  
QY 781 GTCCGCGCTTGTGCCGAAATGAGAAGGATTTCTCTACGCAACCTTGATCCTGAGTA 840





; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-47

Query Match 49.3%; Score 499.4; DB 6; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 1.6e-161;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

OY 2 TGAAGAAGCTATCAAGTGCCTGCGTGCAAGCCGCCGATCTACATGATTGGAGG 61  
DB 2 TGAAGAAGCAATCAAGTAGCTGTGTGCAAGCAGCTCCAGTCTTCTCGACTGGAGC 61  
OY 62 CGACGGTGACAACCAATTGAGTTGATGGAAGACGACAGCTAATATGCTCTGA 121  
DB 62 CCACAGTGACAAGACCGCTGCCCTGATTGAGGAGGACCCGTAACGGCGCAAGCTAA 121  
OY 122 TCGCCTTCCGAACTTGGATTCCAGGCTACCCATGGTTCTTGGCTTGACTCAACAG 181  
DB 122 TCGCCTTCCAGAGACTGATTCAGGCTACCCATGGTTCTTGGCTTGACTCAACAG 181  
OY 182 CATGGCAATGCAATTGTGACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241  
DB 182 CTTGGGGATGCAATTCGTGCGCCGATACCAAGAACTCACTGTCTCGACAGCCCTC 241  
OY 242 AAGCTAAGCGCATTTGATGACGCCAAGCGGTTGGAATCATGTACACCTGGGGATGA 301  
DB 242 AGGCCAAGCGCATGATGAGGCCGCCAGCGCGCGGTATATACGTGCGGCTAGGGTACA 301  
OY 302 GTGAACGGGTGCGTGCAACCTTTACATCAGTCACTGCTTATAGGCGATAATGCTACA 361  
DB 302 GCGAAGCGGTGAGCGGAACCTTACATGAGGGGAGGCTCATTTGACGATAGGGCGAAA 361  
OY 362 CCATTGGGGCCCGGCAAGTTGAAACCTACTTTGTGTAACGTACTTTGTTCGGCGAAG 421  
DB 362 CAGCTGGGCTGCGCCGAAAGCTGAAACCAACCATGTAGAGCGAACCCTTTCGGTGAAG 421  
OY 422 GGGATGTTCACTCGCTAGCGGTTTTCGAGACGCTGCTTGAAGGCTGGGCTTATGCT 481  
DB 422 GCGACGATCATCCCTTTCCACTTTTCGACACACCGTTGGGGGTGCTGGCGACTCTGCT 481  
OY 482 GTTGGAGCACCTTCAACCGCTAACAAATATCGTTGTATGACAAAATGAGAGATTTC 541  
DB 482 GTTGGAAACATTACACTCTTTTCGAATATGCGCTTACGACACAGAGAGAAATATAC 541  
OY 542 ATTGTGCGGCTTGGCGAGCTTTAGCCTTATATGCGGCGAAAGCCCTGGGGCCTG 601  
DB 542 ACTTCGCCGCTGGCCTAGCTTACGATCTACCGTCAAGCGACAGAGAGTCTTGGACACG 601  
OY 602 ATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTGACTAGCGT 661  
DB 602 AAGTAATGTGCGAGCTTCTCGGATCTACGCCGTGGAAGGGCAGTGTCTTCTCGCTT 661  
OY 662 CGTGTGCGCTGTTTCACAATCCATGATGATGCTTGTATCAGATGACGAAAAGCATG 721  
DB 662 CGTGTGCGCTGTTTCACAATCCATGATGATGCTTGTATCAGATGACGAAAAGCATG 721

DB 662 CTTGCGGCTGCTTCGCGCAGAGATGATCGAAATGCTCTGCACTGACGAAAGCAGACACA 721  
OY 722 CGTTGCTTCTGCTGCTGCTGAGACACTACGTATCATAGGGCTGATGGTGTGACTTGG 781  
DB 722 GCCTTCTTCAAGCGCGCGGGGTACTCCCGCATATCGGTCCCGATGGCAGACCTAG 781  
OY 782 TCGCGCCTCTTCCGAAATGAGAGGGTATCTCTACGCAACCTTGATCTGAGTAC 841  
DB 782 CGCGCCCTTGGGCGAAACGAGAGAGTATCTTATGCCACTCTGACCTGCGGCTC 841  
OY 842 GCATCCTTGTCTAAATGCGCGCAGACCCCTGCTGTATATTCCCGTCCGACATTAATC 901  
DB 842 GAATCTATGCAAGACCGCAGCTGATCCAGCGGGCACTACTCCAGACCGACGTCACTC 901  
OY 902 GCTTCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAAATTGAAGGTGATCTTCTG 961  
DB 902 GCGTCTGATCAATCGCATGTCCTCAATCAGCCAGTCGTAGAGGTTGGAAGGGAATACCTG 961  
OY 962 CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014  
DB 962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCGCGGTACGAAAGCGAATGA 1014

RESULT 9  
US-10-440-523-47  
; Sequence 47, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:

; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-47

Query Match 49.3%; Score 499.4; DB 7; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 1.6e-161;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

OY 2 TGAAGAAGCTATCAAGTGCCTGCGTGCAAGCCGCCGATCTACATGATTTGAGG 61  
DB 2 TGAAGAAGCAATCAAGTAGCTGTGTGCAAGCAGCTCCAGTCTTCTCGACTGAGC 61  
OY 62 CGACGGTGACAACCAATTGATGATGGAAGAAGCAGACGATTAATATGCTGCTGA 121  
DB 62 CGACGGTGACAACCAATTGATGATGGAAGAAGCAGACGATTAATATGCTGCTGA 121

```

Db      62  CCACAGTGGACAGAACCGTCGCCCTGATTTGAGAGGACGCCCGTAACGGCGCACGCCCTAA 121
QY      122  TCGCCCTTCCGGAAACTTGATTCAGGGCTACCCATGGTTCTTTGGCTTGAATCACCAG 181
Db      122  TCGCCTTCCAGAGACTGGATTCCAGGCTACCCATGGTTCTTTGGCTTGAATCACCAG 181
QY      182  CATGGCAATGCAATTGTGACGCCAATATACATGAGAACTCATTGGAGTTGGATGGCCCTC 241
Db      182  CCTGGGGATGCAATTGCTGCGCGGATACAGAGAACTCACTGCTCTCGACAGCCCTC 241
QY      242  AAGCTAAGCGCATTTGAGATGACGCCAAGCGGTGGAAATCATGGTCAACCTGGGGATGA 301
Db      242  AGGCCAAGCGCATGAGTGAAGCGCGCCAGCGCGGTATATACGTGCGCTAAGGTACA 301
QY      302  GTGAACGGGTGCGTGACACCTTTACATCAGTGTGTTGTAAGCGATAGGCGATATGTTACA 361
Db      302  GCGAACGGGTGAGCGGAACCTTACATGCGGAGTGGCTCATTTGACGATTAAGGGCGAAA 361
QY      422  GGGATGTTTATCGCTAGCGGTTTTCGAGACGTCTGTTGAAGGCTGGGTGCTTATGCT 481
Db      422  GCGACGGATCATCCCTTTCACATTTGACACACCGTTGGGGGTGCTGGGCGACTGTGCT 481
QY      482  GTTGGGAGCACCTTCAACCGCTTAACAAATACGCTTTGTATGCACAAAATGAGAAGATTTC 541
Db      482  GTTGGGAACACTTACAACTCTTTTCGAATATGCGCTCTACGACAGAGAGAGAAATAC 541
QY      542  ATTGTGCGGCTTGCGCGGCTTTAGCCTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
Db      542  ACTTGGCGGCTTGCGGCTTACCTTACATCTACCGCTCAAGCGACAGAAAGCTTGGACCA 601
QY      602  ATGTCAATGTAGCGGCTCTGCAATCTATGCCGTTGAAGGCAATGCTTCTGTTAGCGT 661
Db      602  AAGTAAATGTGCGAGCTTCTCGATCTACGCCGTGAAAGGCGAGTGTGTTGTTCTGCTT 661
QY      662  CGTGTGCGGCTGTTTCAATCCATGATGATGATGATGATGATGATGATGATGATGATG 721
Db      662  CTTGCGGCTGCTCTCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATG 721
QY      722  CGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781
Db      722  GCGTCTTCTGAGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
QY      782  TCGGCGCTCTTGGCGGAAATGAGAGGGTATTTCTTACGCAAACTTGTATCTGAGTAC 841
Db      782  CGCGCGCTTGGCGGAAATGAGAGGGTATTTCTTACGCAAACTTGTATCTGAGTAC 841
QY      842  GCATCCTTGTAAATGCGGCGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901
Db      842  GAATCTATGCAAAAGACCGCAGCTGATCCAGCGCGGCACTACAGACCAAGACGTCAC 901
QY      902  GCTTGTATATAGATCGCAGCCCTAAATTAACCGTAGTTGAATGAAGTGTATCTGCTC 961
Db      902  GCGTGTATATAGATCGCAGCTGATCCAGCTGATCCAGCTGATCCAGCTGATCCAGCT 961
QY      962  CTTACGCTTTGGGTAAGCGCTGAGACGGGTGCGCAACTCGAAGAAATTTGA 1014
Db      962  CATCGGCGCAAGGCTTTGAAGTTGAGGCGGCGCGCGGTACGAAGCGCATTTGA 1014

```

RESULT 10  
US-10-440-503-47

; Sequence 47, Application US/10440503  
; Publication No. US20040038419A1

; GENERAL INFORMATION:

; APPLICANT: Weiner, David Paul

; APPLICANT: Chaplin, Jennifer Ann

; APPLICANT: Chi, Ellen

; APPLICANT: Milan, Aileen

```

; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-47

```

Query Match 49.3%; Score 499.4; DB 7; Length 1014;  
Best Local Similarity 68.3%; Pred. No. 1.6e-161;  
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

```

QY      2  TGAAGAAGCTATCAAGTTCGCGCTGCGTCAAGCCGCCGATCTACATGATTTGAGG 61
Db      2  TGAAGAAGCAATCAAGTAGCCTGTGTGCAAGCAGCTCCAGTCTTCTGACCTGAGC 61
QY      62  CGACGGTGAACAAACCATTTGATGATGGAAGAAGACGACGTAATATGCTGCTGA 121
Db      62  CCACAGTGGACAGAACCGTCGCCCTGATTTGAGAGGACGCCCGTAACGGCGCACGCCCTAA 121
QY      122  TCGCCTTCCGAAACTTGATTTCAAGCTACCCATGTTCTTTGGCTTGAATCACCAG 181
Db      122  TCGCCTTCCAGAGACTTGATTTCAAGCTACCCATGTTCTTTGGCTTGAATCACCAG 181
QY      182  CATGGCAATGCAATTGTAGCGCAATACATGAGAACTCATTTGAGTGTGATGGCCCTC 241
Db      182  CCTGGGATGCAATTCGTGCGCGCATACAGAGAACTCACTGTGCTTGAAGCCCTC 241
QY      242  AAGCTAAGCGCATTTAGATGACGCCAAGCGGTGGAAATCATGATGATGATGATGATG 301
Db      242  AGGCCAAGCGCATGATGAGCGCGCGCGCGGTATATACGTGCGCTAAGGTACA 301
QY      302  GTGAACGGGTGCGTGCAACCTTTACATCAGTGTGTTGATAGGCGAATATGTTGACA 361
Db      302  GCGAACGGGTGAGCGGAACTCTTACATGAGGCGAGTGTGATGATGATGATGATGATG 361
QY      362  CCATTGGGGCGCGGAAAGTTGAAACCTTCTTGTGAACGTAATTTGTCGGCGAAG 421
Db      362  CAGCTGGGCTGCGCGGAAAGCTGAAACCAACCATGTAGAGCGAACCTCTTGGTGAAG 421
QY      422  GGGATGTTTATCGCTAGCGGTTTTCGAGACGTCTGTTGAAAGGCTGGGTGCTTATGCT 481
Db      422  GCGACGATCATCCCTTTCACATTTGACACACCGTTGGGGGTGCTGGGCGGACTCTGCT 481
QY      482  GTTGGAGCACCTTCAACCGCTTAACAAATACGCTTTGTATGACAAATGAGAAGATTTC 541
Db      482  GTTGGGAACACTTACAACTCTTTGAAATATGCTCTTACGACAGAAAGAGAAATAC 541
QY      542  ATTGTGCGGCTTGCGCGGCTTTAGCCTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
Db      542  ACTTGGCGGCTTGCGCGGCTTACGATCTACCGTCAAGCGACAGAAAGTCTTGAACAG 601
QY      602  ATGTCAATGTAGCGGCTCTGCAATCTATGCGGTTGAAGGCGAATGCTTCTGATAGCGCT 661
Db      602  AAGTAAATGTGCGAGCTTCTCGATCTACGCCGTGAAAGGCGAGTGTGTTGTTCTGCTT 661
QY      662  CGTGTGCGGCTGTTTCAATCCATGATGATGATGATGATGATGATGATGATGATGATG 721
Db      662  CTTGCGGCTGCTCTCGCGCAGATGATGATGATGATGATGATGATGATGATGATGATG 721

```



QY 722 CGTTGCTTCTGGCTGTGTGGACACTCACGTATCATAGGGCCTGATGTGTGACTTGG 781  
Db 722 GCCTTCTTCAGGCCGCGCGGGGTACTCCCGCATTCGGTCCGATGGCAGCACTAG 781  
QY 782 TCGCGCCTCTTGCCGAAATGAGAGGGTATCTCTACGCAAACTTGATCTGAGTAC 841  
Db 782 CGCGCCCTTGGCGCAAAACGAGGAAGTATCTCTATGCCACTGTGACCCTGCCGCTC 841  
QY 842 GCATCTTGCTAAATGCGGCAGACCCTGTGTCATTATTCCGTCGCCAGACTTACTC 901  
Db 842 GAATCTATGCAAGAACCAGAGCTGATCCAGCCGGGCACACTACTCCAGACCAGCTACTC 901  
QY 902 GCTTGCTATAGATCGCAGCCCTAATTACCGTAGTTGAAATTGAAGTGATCTTCGTC 961  
Db 902 GGCTGCTGATCAATCGCAGTGCCTAATCAGCCAGTCTGTAGAGTTGGAAGGGAATACTCG 961  
QY 962 CTTAGCCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGCA 1014  
Db 962 CATCGGCCCAAGGCTTTGAAAGTTGAGGCGGCGCTCCGGGTACGAAGGCCGATTGA 1014

## RESULT 11

```

US-10-461-925-47
; Sequence 47, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461, 925
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389, 317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392, 944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-47

```

Query Match	49.3%	Score 499.4;	DB 7;	Length 1014;
Best Local Similarity	68.3%;	Pred. No. 1.6e-161;		
Matches 692; Conservative	0;	Mismatches 321;	Indels 0;	Gaps 0;

[illegible]

Db 302 GCCAACCGCGTGAGCGGAAACCCCTCTACATGGGGCAGTGGCTCATTTGACGATTAAGGCCGAAA 361

Qy 362 CCATTGGGGCCCCGGCGAAAGTTGAAACCTACTTTTGTGTGAACGTACTTTGTTGGCGAAG 421

Db 362 CAGCTGGGCTGCGCCGAAAGCTGAACCAACCCATGTAGAGCGAAACCTCTTCGTGAAG 421

Qy 422 GGGATGCTTCATCGCTAGCGGTTTTGAGACGTCTGTTGGAAGGCTGGGTGCTTATGCT 481

Db 422 GCGACGGATCATCCCTTTCCACTTCGACACACCGCTGGGGGTGCTGGCGGACTCTGCT 481

Qy 482 GTTGGGAGCACCTTCAACCGCTAACAAATAACGCTTGTATGCACAAATGAGAAGATTTC 541

Db 482 GTTGGGAACACTTACAACCTCTTTGAAATATGCGCTCTACGCACAGAACGAGAAATAC 541

Qy 542 ATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGGCTTG 601

Db 542 ACTTCGCGCCTGCGCTAGCTTCAGCATCTAACGTCAGCGACAGAAAGTCTTGAACAG 601

Qy 602 ATGTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGCGAATGCTTCGTACTAGCGT 661

Db 602 AAGTAATGTGCGCAGCTTCTCGGATCTACGCCGTGAAGGCGAGTGTTTGTTCTCGCTT 661

Qy 662 CGTGTGCGCTCGTTTACAAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATG 721

Db 662 CTGCGCGCTCGTCTCGCCAGAGATGATCGAAATGCTCTGCATGACGAAAGCAAGCACA 721

Qy 722 CGTTGCTTCTGCGTGTGTGACACTCACGATCATAGGCGCTGATGTGTGACTTGG 781

Db 722 GCCTTCTTCAAGCGCGCGGGTACTCCGCATTATCGTCCCGATGGCAGCAGACTAG 781

Qy 782 TCGCGCCTCTTGCCGAAATGAAGAGGTATTCTCTACGCAAACTTGATCCTGAGATAC 841

Db 782 CGCGCCCTTGCGCGAATAACGAGAGAGTATTCTCTATGCCACTCTGGAACCTGCGGCTC 841

Qy 842 GCATCCTTGTAAATGCGCGGACAGCCCTGCTGTGCTATTATCCGTCGCCGACATTACTC 901

Db 842 GAATCTATGCAAAAGACGCACTGATCCAGCCGGGCACTACTCCAGACAGACGTCACCTC 901

Qy 902 GCTTGCTAATAGATCGCAGCCCTAAATTACCGGTAGTTGAAATTGAAGTGATCTTCGTC 961

Db 902 GGCTGTGATCAATCGCAGTGCATACAGCCAGTGTAGAGGTTGGAAGGAATAATACCTG 961

Qy 962 CTTAAGCTTTGGTAAAGCGTCTGAGACGGGTGCGCACTCGAAGAAATTGA 1014

Db 962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCCCCCGGCTACGAAAGCGGATTGA 1014

## RESULT 12

US-10-146-772-57  
 ; Sequence 57, Application US/10146772  
 ; Publication No. US20030124698A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Short, Jay  
 ; APPLICANT: Weiner, David  
 ; APPLICANT: Chaplin, Jennifer  
 ; APPLICANT: Chi, Ellen  
 ; APPLICANT: Milan, Aileen  
 ; APPLICANT: Desantis, Grace  
 ; APPLICANT: Madden, Mark  
 ; APPLICANT: Burk, Mark  
 ; TITLE OF INVENTION: Nitriclases  
 ; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
 ; CURRENT APPLICATION NUMBER: US/10/146,772  
 ; CURRENT FILING DATE: 2002-05-15  
 ; PRIOR APPLICATION NUMBER: US 60/309,006  
 ; PRIOR FILING DATE: 2001-07-30  
 ; PRIOR APPLICATION NUMBER: US 60/351,336  
 ; PRIOR FILING DATE: 2002-01-22  
 ; PRIOR APPLICATION NUMBER: US 60/300,189  
 ; PRIOR FILING DATE: 2001-06-21  
 ; PRIOR APPLICATION NUMBER: US 09/751,299  
 ; PRIOR FILING DATE: 2000-12-28



; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 57  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-57

Query Match 49.1%; Score 497.8; DB 6; Length 1014;  
Best Local Similarity 68.2%; Pred. No. 5.9e-161;  
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGGTCCGCTGCGTCAAGCCGCCGATCTACATGATTTGAGG 61  
DB 2 TGAAGAAGCAATCAAGTAGCCTGTGTGCAAGCAGCTCCAGTCTTTCTGACCTGAGC 61  
QY 62 CGACGGTGACAAAACCATTTGAGTTGATGGAAGAACAGACAGCATATATGCTCGTGA 121  
DB 62 CCACAGTGACAAAGACCGTCCCTGATTGAGAGGAGCCCGTAAACGGCCACGCTAA 121  
QY 122 TCGCCTTTCCGAAACTTGAGTTCCAGGCTACCCATGGTTTCTTGGCTTACTCACCAG 181  
DB 122 TCGCCTTTCCAGAGACCTGAGATTCAGGCTACCCATGGTTCTTTGGCTGACTCACCAG 181  
QY 182 CATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241  
DB 182 CCTGGGGGATGCAATTCGTGCGCCGATACCAAGAACTCACTGTTCTTCGACGCCCTC 241  
QY 242 AAGCTAAGCGCATTTCAATGACGACCAAGCGGTTGGAAATCATGTGTACCCCTGGGATGA 301  
DB 242 AGGCCAAGCGCATCAGTAGGCGCCGACGCGCCGGTATATACGTGCGCTAGGGGTACA 301  
QY 302 GTGAACGGGTGCGTGACACCTTTACATCAGTCACTGTTGTAACGTACTTTGTCGGGAG 361  
DB 302 GCGAACGGGTGACGGAACCTCTACATGGGCACTGCTCATTTGACGATTAAGGCGAAA 361  
QY 362 CCATGGGGCCCCGCGAAAGTTGAAACCTACTTTGTTGAACGTACTTTGTCGGGAG 421  
DB 362 CAGTGGGCTGCGCGGAAAGCTGAACCAACCATGTAGAGCAACCTCTTCGTGTAAG 421  
QY 422 GGGATGGTTATCGCTAGCGGTTTCGAGACGTCTGTGGAAGGCTGGGTGCTTATGCT 481  
DB 422 GCGACGATCATCCCTTTCCACTTTGACACACCGTTGGGGGTGCTGGGCGACTGTCT 481  
QY 482 GTTGGAGACCTTCAACCGCTAACAAATAAGCTTTGTATGACAAATGAAGAGATTC 541  
DB 482 GTTGGAGACCTTCAACCTCTTTGAAATATGCGCTCTACGACAGAAAGAAATAC 541  
QY 542 ATTGTGGGCTTTGGCCGAGCTTTAGCCCTTATCCTAATGCGGCGAAAGCCCTGGGCTG 601  
DB 542 ACTTGGCGGCTGGCTAGCTTCAAGCATCTACCGTCAAGCGAGAGAGTCTTGAACAG 601  
QY 602 ATGTCAATGAGCGGCTCTCGAATCTATGCGCTTGAAGGGCAATGCTTCTACTAGCGT 661  
DB 602 AAGTAATGTCGAGCTTCTCGGATCTACCGCGGTGAAGGGCAATGTTTGTCTCGCTT 661  
QY 662 CGTGTGCGCTGTTTCACAATCCATGATGATGCTTTGTATCAGATGACGAAAAGCATG 721  
DB 662 CCTGCGGCTGCTGCTGCGCAGAGATGATGAAATGCTCTGCACTGACGAAAAGCAAGACA 721  
QY 722 CGTTGCTTGGCTGCTGCTGCTGACACTACGATCATAGGGCTGATGGTGTGACTTGG 781  
DB 722 GCCTTCTTCAAGGCGCGCGGGGTACTCCGCAATATCGGTCCCGATGGCAGGCACTAG 781  
QY 782 TCGGCGCTCTTGGCGAAATGAAGAGGATTTCTTACGCAAACTTGATCTGAGACTAC 841  
DB 782 CGCGCGCTTGGCGGAAAGAGAGGATTTCTTATGCACTCTGCACTGCGGCTC 841

QY 842 GCATCCTTGCTMAAATGCGCGACAGACCTTGTGTCATTATTTCCCGTCCGACATTA 901  
DB 842 GAATCTATGCAAGAACCGCAGCTGATTCAGCGGGCACTACTCCAGACGACGTCACTC 901  
QY 902 GCTTGCTAATAGATGACGAGCCCTAAATTACCGGTAGTTGAAATTGAAGTGA 961  
DB 902 GGCTGTGATCAATCGAGTGCCCAATCAGCCAGTGTAGAGTTGGACGGGAAATACCTG 961  
QY 962 CTACGCTTTGGTTAAAGCTGTGAGACGGGTGCGCACTCGAAGAAATTTGA 1014  
DB 962 CATCGGCCCAAGCTTTGAAGTTGAGCGGCGCCCGGGTACGAGGCGGATTGA 1014

RESULT 13  
US-10-241-742-57  
; Sequence 57, Application US/10241742  
; Publication No. US20040002147A1

; GENERAL INFORMATION:

; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark

; TITLE OF INVENTION: Nitriases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; CURRENT FILING DATE: 2002-09-09

; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30

; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21

; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07

; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 57  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown

; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-57

Query Match 49.1%; Score 497.8; DB 6; Length 1014;  
Best Local Similarity 68.2%; Pred. No. 5.9e-161;  
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGGTCCGCTGCGTCAAGCCGCCGATCTACATGATTTGAGG 61  
DB 2 TGAAGAAGCAATCAAGTAGCCTGTGTGCAAGCAGCTCCAGTCTTTCTGACCTGAGC 61  
QY 62 CGACGGTGACAAAACCATTTGAGTTGATGGAAGAACAGACAGCATATATGCTCGTGA 121  
DB 62 CCACAGTGACAAAGACCGTCCCTGATTTGAGAGGAGCCCGTAAACGGCGCAAGCTAA 121  
QY 122 TCGCCTTTCCGAAACTTGAGTTCCAGGCTACCCATGGTTTCTTGGCTTACTCACCAG 181  
DB 122 TCGCCTTTCCAGAGACCTGATTTCAAGCTACCCATGGTTCTTGGCTGACTCACCAG 181  
QY 182 CATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTC 241  
DB 182 CGCGCGCTTGGCGGAAAGAGGATTTCTTATGCACTCTGCACTGCGGCTC 241

182 CTTGGGGGATGCAATTCTGCGCCGATACCAAGAGAACTACTGGTCTCGACAGCCCTC 241  
QY 242 AAGCTAAGCGCATTTAGATGACCCAAAGCGGTTGGAAATCATGTCAACCTGGGAGTGA 301  
Db 242 AGGCCAAGCGCATGAGTGGCGCCAGCGCGCGGTATATAGCTGCGCTAGGATACA 301  
QY 302 GTGAACGGGTGCGTGACACCTTTACATCAGTCAGTGGTTCATAGGCGATAATGATGACA 361  
Db 302 GCGAACGCGTGAGCGGAACCTCTACATGGGGCAGTGGCTCATGACGATAAGGCGGAAA 361  
QY 362 CCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAG 421  
Db 362 CAGCTGGGCTGCGCCGAAAGCTGAAACCAACCATGTAGAGCGAACCCCTCTTCGGTGAAG 421  
QY 422 GGGATGGTTCATCGCTAGCGGTTTTTCGAGACGCTGTGTTGAAGGCTGGGTGCTTATGCT 481  
Db 422 GCGACGGATCATCCCTTTCCACTTTTCACACACACCGTTGGGGGTGCTGGCGGACTGCT 481  
QY 482 GTTGGGAGCACCTTCAACCGCTAACAAATATAGCTTTGTATGCAAAATGAAGAGATTC 541  
Db 482 GTTGGGAACACTTACAAACCTCTTTGAAATATAGCGCTTACGACACAGAACGAAATAC 541  
QY 542 ATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCGTG 601  
Db 542 ACTTCGCGCGCTGCGCTAGCTTACGATCTACCGTCAAGCGACAGAAGTCTTGACAG 601  
QY 602 ATGTCAATGTAGCGGCTCTGCAATCTATGCGGTTGAAGGCAATGCTTCGTACTAGCGT 661  
Db 602 AAGTAAATGTGCGAGCTTCTCGGATCTACGCGGTGAAGGCGAGTGTTCGTTCTCGCTT 661  
QY 662 CGTGTGCGCTGTTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATG 721  
Db 662 CCGTGGCGCTGCTGCTGCGAGATGATGAAATGCTGTCGACGAAAGCAAGCACA 721  
QY 722 CGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781  
Db 722 GCGTCTTCAAGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
QY 782 TCGCGCTCTTCCGAAATGAAGAGGATATTCTTACGCAACCTTGATCCTGAGTAC 841  
Db 782 CGCGCCCTTGGCGGAAACGAGAGATATTCTTATGCCACTTGAGCCCTGCGCTC 841  
QY 842 GCATCCTTGTAAATGGCGGAGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901  
Db 842 GAATCTATGCAAAAGACCGGAGCTGATCCAGCCGGGCACTACTCCAGACCAGACGTACTC 901  
QY 902 GCTTGCTAATAGATCGCAGCCCTAAATTACCGGTTAGTTGAATTGAAGGTGATCTTCGTC 961  
Db 902 GCGTGTGATCAATCGCAGTGCATCAGCCAGTCTGTAGAGGTTGACGGGAAATACCTG 961  
QY 962 CTTACGCTTTGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAAGAAATTGA 1014  
Db 962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCCCCGGGTAGGAGGCGATTGA 1014

RESULT 14  
US-10-440-523-57  
; Sequence 57, Application US/10440523  
; Publication No. US20040014195A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitriases  
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/440,523  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US/10/146,772

; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 57  
; LENGTH: 1014  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-523-57  
Query Match 49.1%; Score 497.8; DB 7; Length 1014;  
Best Local Similarity 68.2%; Pred. No. 5.9e-161;  
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;  
QY 2 TGAAGAAGCTATCAAGTGCCTGCGTGAAGCCCGCATCATGATGATTTGAGG 61  
Db 2 TGAAGAAGCAATCAAGTAGCCTGTGTGAAGCAGCTCCAGTCTTCTCGACTGAGC 61  
QY 62 CGACGGTGACAAACCATTGAGTTGATGGAAGAGCAGCAGCTAATATGCTGCTGA 121  
Db 62 CCACAGTGACAAAGCCGTCGCCCTGATTGAGAGGAGCCCGTAAACGGCGCACGCTAA 121  
QY 122 TCGCCTTCCGAAACTTGGATTCCAGGCTTACCAATGCTTTCTTGGCTTGAACCA 181  
Db 122 TCGCCTTCCAGAGACTGGAATCCAGGCTTACCAATGCTTTCTTGGCTTGAACCA 181  
QY 182 CATGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGTTGATGCGCTC 241  
Db 182 CTTGGGGATGCAATTCGTGCGCCGATACCAAGAACTCACTGCTGCTGACAGCCCTC 241  
QY 242 AAGCTAAGCGCATTTAGATGACGCCAAGCGGTTGGAAATCATGATGATGATGATGATG 301  
Db 242 AGGCCAAGCGCATGAGTGAAGCGCCGAGCGCGGTATATACGTGCGGTAGGGTACA 301  
QY 302 GTGAACGGGTGCGTGACACCTTTACATCAGTCAGTGGTTCATAGGCGATAATGATGACA 361  
Db 302 GCGAACGCGTGAGCGGAAACCTCTACATGGGGCAGTGGCTCATGACGATAAGGCGGAAA 361  
QY 362 CCATTGGGGCCCCGGGAAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAAG 421  
Db 362 CAGCTGGGCTGCGCGGAAAGCTGAAACCAACCATGTAGAGCGAACCCCTTTCGGTGAAG 421  
QY 422 GGGATGGTTCATCGCTAGCGGTTTTTCGAGACGCTGTGGAAGGCTGGGCTTATGCT 481  
Db 422 GCGACGATCATCCCTTTCCACTTTTCGACACACCGTTGGGGGTGCTGGCGGACTCTGCT 481  
QY 482 GTTGGAGACACTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGAGATTTC 541  
Db 482 GTTGGGAACACTTACAACTCTTTGAAATATGCGCTCTACGACACAGAACGAAATAC 541  
QY 542 ATTGTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCGCTG 601  
Db 542 ACTTCGCGCGCTGCGCTAGCTTACGATCTACCGTCAAGCGACAGAAAGTCTTGAGACAG 601  
QY 602 ATGTCAATGTAGCGGCTCTGCAATCTATGCGGTTGAAGGCGAATGCTTCTGTAAGCGT 661  
Db 602 AAGTAAATGTGCGAGCTTCTCGGATCTACGCGGTGAAGGCGAGTGTTCGCTTCGCTT 661  
QY 662 CGTGTGCGCTGTTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATG 721  
Db 662 CGTGTGCGCTGTTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATG 721

```

Db      662 CCTGCGCGCTGCTCTCGCCAGAGATGATCGAAATGCTCTGCACTGACGAAAGCAAGCACA 721
QY      722 CGTTGCTTCTGAGCTGTGTGACACTACGTATCATAGGCGCTGATGGTGTGACTTGG 781
Db      722 GCCTTCTTCAAGCCGGCGGGGTACTCCCGATTATCGGTCCCGATGGCAGCAGCCTAG 781
QY      782 TCGCGCCTTGTGCGGAAATGAAGAGGGTATTCTCTACGCAAACTTGATCCTGAGTAC 841
Db      782 CGCGCCCTTGGGGGAAAAAGAGGAAGTATTCTCTATGCCACTTGGAACCTGCCGCTC 841
QY      842 GCATCCTTGTAAATGGCGGACAGCCCTGTGTCATTATTCCCGTCCGACATTACTC 901
Db      842 GAATCTATGCAAGACCGCAGCTGATCCAGCGGGCACTACTCCAGACGAGCTCACTC 901
QY      902 GCTTGCTAATAGATCCGAGCCCTAATTAACCGGTAGTTGAATGGAAGTGTATCTTCTC 961
Db      902 GGCTGCTGATCAATCGCAGTCCCAATCAGCCAGTCGTAGAGTTGAGCGGAAATACCTG 961
QY      962 CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014
Db      962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCGGCGGTACGAGGCGAATTGA 1014

```

```

RESULT 15
US-10-440-503-57
; Sequence 57, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE OF INVENTION: NITRILES AND/OR CYANIDE
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-440-503-57

```

```

Query Match      49.1%; Score 497.8; DB 7; Length 1014;
Best Local Similarity 68.2%; Pred. No. 5.9e-161;
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

```

```

QY      2 TGAAGAAGCTATCAAGTGCCTGCGTGAAGCGCCCGATCTACATGATTTGGAGG 61
Db      2 TGAAGAAGCAATCAAGTAGCCTGTGTGAAGCAGCTCCAGTCTTTCTGACCTGGACG 61
QY      62 CGACGCTGACAAACCATTTGATGTATGGAAGACAGACGTAATAATGCTGCTGA 121
Db      62 CCACAGTGACAGACCGCTCGCCTGATTTAGAGGACGCGGTAAACGCGCACGCTAA 121
QY      122 TCGCCTTTCCGGAACCTTGAATTCAGGCTACCAATGTTCTTTGCTTGACTCAGCAG 181
Db      122 TCGCCTTTCCAGAGACCTGATTCAGGCTACCAATGTTCTTTGCTTGACTCAGCAG 181
QY      182 CATGGCAATGCAATTTGTACGCCCAATACCATGAGAACTCATTTGAGTTGAGTGGCCCTC 241
Db      182 CTTGGGGGATGCAATTCTGTGCGCCGATACCAAGAACTCACTGTCTCGACAGCCCTC 241

```

```

QY      242 AAGCTAAGCGATTTCAAGTACAGCCAAAGCGGTGGAAATCATGTGTACCCCTGGGATGA 301
Db      242 AGGCCAAGCGCATCAGTGAGGCGCCCGCCCGCGGTATATAGTCGCGTAGGGTACA 301
QY      302 GTGAACGGGTGCGTGGACACCTTTACATCAGTCAAGTGTTCATAGCGGATAATGTGACA 361
Db      302 GCGAAGCGGTGAGCGGAAACCTCTTACATGCGGCGAGTGGCTCATTTGACGATAAGGCGAAA 361
QY      362 CCATTGGGGCCCGCGGAAAGTTGAAACCTTCTTTGTTGAACGTAATTTGTTCGGCGAAG 421
Db      362 CAGCTGGGCTGCGCGGAAAGCTGAAACCAACCATGTAGAGCGAACCCTCTTCGGTGAAG 421
QY      422 GGGATGTTTCAATCGCTAGCGGTTTTCAGACGCTCTGTTGAAAGGCTGGGCTTATGCT 481
Db      422 GCGACGGATCATCCCTTTCACATTTGACACACCGTTGGGGGTGCTGGGCGGACTCTGCT 481
QY      482 GTTGGGAGCACCTTCAACCGCTAACAAATATAGCTTTGTATGACACAAATGAGAAGATTC 541
Db      482 GTTGGGAACACTTACAACTCTTTGAAATATGCGCTTACGACACAGAAAGAAATAC 541
QY      542 ATGTGCGGCTTGGCGGAGCTTTAGCCTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
Db      542 ACTTGGCGGCTTGGCGGAGCTTTAGCCTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
QY      602 ATGTCAATGTAGCGGCTTTCGAATCTATGCGGTTGAAGGCAATGCTTGTACTAGCGT 661
Db      602 AAGTAAATGTGCGAGCTTCTCGAATCTAGCGCGTGAAGGGCAGTGTCTTGTCTCGCTT 661
QY      662 CGTGGCGGCTGTTTCAACATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATG 721
Db      662 CTTGCGGCTGCTCTGCGGAGAGATGATGAAATGCTGTGCACTGACGAAAGCAGACA 721
QY      722 CGTTGCTTCTGCGCTGTGTGACACTCAAGTATCATATAGGGCTGTGTTGTTGACTTGG 781
Db      722 GCCTTCTTCAAGCGCGGCGGGGTACTCCCGCATATCGGTCCCGATGGCAGCGACTAG 781
QY      782 TCGCGCCTTGTGCGGAAATGAAGGGTATTCTCTACGCAAACTTGATCCTGAGTAC 841
Db      782 CGCGCCCTTGGGGGAAAAAGAGGAAGTATTCTCTATGCCACTTGGAACCTGCGGCTC 841
QY      842 GCATCCTTGTAAATGGCGGACAGCCCTGCTGTCATTATTCCTCCGACATTACTC 901
Db      842 GAATCTATGCAAGACCGCAGCTGATTCAGCCGGGCACTACTCCAGACGAGCTCACTC 901
QY      902 GCTTGCTAATAGATCGCAGCCCTAATTAACCGGTAGTTGAATGGAAGTGTATCTTCTC 961
Db      902 GGCTGCTGATCAATCGCAGTCCCAATCAGCCAGTGTAGAGGTTGACGCGGAAATACCTG 961
QY      962 CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTGA 1014
Db      962 CATCGGCCCAAGGCTTTGAAGTTGAGGCGGCGGCGGTACGAGGCGAATTGA 1014

```

```

Search completed: April 27, 2006, 00:01:30
Job time : 979.006 secs

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:27:34 ; Search time 609.88 Seconds  
(without alignments)  
6752.353 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1014

Sequence: 1 atggaagaagctatcaaggt.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA\_New:\*

1: /SIDS5/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
2: /SIDS5/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /SIDS5/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
4: /SIDS5/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
5: /SIDS5/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
6: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq1:\*  
7: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
8: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq1:\*  
9: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq2:\*  
10: /SIDS5/ptodata/2/pubpna/US10\_NEW\_PUB.seq3:\*  
11: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
12: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq2:\*  
13: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq3:\*  
14: /SIDS5/ptodata/2/pubpna/US11\_NEW\_PUB.seq4:\*  
15: /SIDS5/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248.6	24.5	1071	9	US-10-537-075-6 Sequence 6, Appli
2	219.6	21.7	1110	9	US-10-919-182-15 Sequence 15, Appl
3	219.6	21.7	1110	9	US-10-919-182-17 Sequence 17, Appl
4	218	21.5	1110	9	US-10-919-182-3 Sequence 3, Appli
5	218	21.5	1110	9	US-10-919-182-5 Sequence 5, Appli
6	218	21.5	1110	9	US-10-919-182-11 Sequence 11, Appl
7	216.4	21.3	1110	9	US-10-919-182-7 Sequence 7, Appli
8	216.4	21.3	1110	9	US-10-919-182-13 Sequence 13, Appli
9	83	8.2	1415	11	US-11-096-568A-6953 Sequence 6953, Ap
10	81.6	8.0	1706	11	US-11-096-568A-20685 Sequence 20685, A
11	70.4	6.9	597	9	US-10-932-182A-166788 Sequence 166788,
12	70.4	6.9	597	9	US-10-932-182A-166788 Sequence 166788,
13	34.6	3.4	548	6	US-09-925-065A-628387 Sequence 628387,
14	34	3.4	548	6	US-09-925-065A-425677 Sequence 425677,
15	34	3.4	563	10	US-10-301-480-489364 Sequence 489364,
16	34	3.4	563	10	US-10-301-480-1102773 Sequence 1102773,
17	33.6	3.3	579	9	US-10-301-480-56141 Sequence 56141, A
18	33.6	3.3	579	10	US-10-301-480-669550 Sequence 669550,

19	33.2	3.3	3815	14	US-11-099-855-1	Sequence 1, Appli
20	33	3.3	42379	9	US-10-330-773-637	Sequence 637, App
21	32.8	3.2	609	9	US-10-301-480-75064	Sequence 75064, A
22	32.8	3.2	609	10	US-10-301-480-688473	Sequence 688473,
23	32.8	3.2	2842	10	US-10-644-807-52	Sequence 52, Appl
24	32.6	3.2	336	8	US-10-623-155-286	Sequence 286, App
25	32.4	3.2	4620	14	US-11-136-527-3028	Sequence 3028, Ap
26	32	3.2	1218	8	US-10-750-185-64760	Sequence 64760, A
27	32	3.2	1218	8	US-10-750-623-64760	Sequence 64760, A
28	31.6	3.1	574	6	US-09-925-065A-368038	Sequence 368038,
29	31.6	3.1	576	10	US-10-301-480-438304	Sequence 438304,
30	31.6	3.1	576	10	US-10-301-480-1051713	Sequence 1051713,
31	31.6	3.1	984	10	US-10-301-480-531203	Sequence 531203,
32	31.6	3.1	984	10	US-10-301-480-1144612	Sequence 1144612,
33	31.4	3.1	621	10	US-10-301-480-475097	Sequence 475097,
34	31.4	3.1	621	10	US-10-301-480-1088506	Sequence 1088506,
35	31.4	3.1	622	6	US-09-925-065A-408575	Sequence 408575,
36	31.4	3.1	1471	11	US-11-183-327-1	Sequence 1, Appli
37	31.4	3.1	1471	11	US-11-183-327-24	Sequence 24, Appl
38	31.4	3.1	1702	11	US-11-072-512-919	Sequence 919, App
39	31.4	3.1	1905	11	US-11-072-512-1070	Sequence 1070, Ap
40	31.4	3.1	3289	8	US-10-750-185-31224	Sequence 31224, A
41	31.4	3.1	3289	8	US-10-750-623-31224	Sequence 31224, A
42	31.4	3.1	3612	6	US-09-925-065A-702369	Sequence 702369,
43	31.4	3.1	4053	11	US-11-072-512-539	Sequence 539, App
44	31.2	3.1	513	6	US-09-925-065A-187799	Sequence 187799,
45	31.2	3.1	513	6	US-09-925-065A-187800	Sequence 187800,

ALIGNMENTS

RESULT 1  
US-10-537-075-6  
; Sequence 6, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zelinski, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 1071  
; TYPE: DNA  
; ORGANISM: Alcaligenes faecalis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1068)  
; OTHER INFORMATION: coding for nitrilase  
US-10-537-075-6

Query Match 24.5%; Score 248.6; DB 9; Length 1071;  
Best Local Similarity 54.6%; Pred. No. 3.8e-70;  
Matches 497; Conservative 0; Mismatches 414; Indels 0; Gaps 0;

QY	4	AAAGAAGCTATCAAGTCGCTGCTGCAAGCCGCCGATCTACATGATTTGAGCG	63
DB	10	AGAAATCTGTCGGGACGCCCGTACAGGCCGCTTCCCACTACGATCTGGCAACG	69
QY	64	ACGGTGACAAACCATTTGATTTGATGAAGAAGCAGCAGTAATAATGCTGCTGATC	123
DB	70	GGTGTGATAAACCATTTGAGCTGCTGCTGACAGCCCGCATGAGGCTGACTGATC	129
QY	124	GCCTTCCGAAACTTGATTTCCAGGCTACCATGCTTTCTTGGCTTGACTCACAGCA	183



Db 130 GTGTTGGTGAACCTGGCTGCCCCGAGATATCCCTTCCACGCTCGGCTGGCGCACCGGCC 189

QY 184 TGGGCAATGCATTTGTACGCCAATACCATGAGAACTCATTTGGAAGTTGGATGGCCCTCA 243

Db 190 TGGTCGCTGAATAACAGTGGCCGCTACTATGCCAACTCGCTCTCGCTGGACAGTGACAG 249

QY 244 GCTAAGCGCATTTCAAGATGACGCCAAGCGGTTGGGAATCATGTGTCACCCTGGGGATGAT 303

Db 250 TTTCACGCGATTGCCAGCGCCGACCGAACCTTGGGTATTTTCATCGCACTGGGTTATAGC 309

QY 304 GAACGGGTGGTGGCACCCCTTTACATCAGTCAGTGGTTCATAGGCCATATGTTGACACC 363

Db 310 GAGCGCAGCGCGGACGCTTTACTGGGCCAATGCTGATGACGACCAAGGCGGAGATG 369

QY 364 ATTGGGGCCCCGGGAAAGTTGAAACCTACTTTTGTGAACTGACTTTGTTCCGCGAAGG 423

Db 370 CTGTGTGCGCTGCAAACTCAAAACCACGCATGTAGAGCCGACCGTATTTGTGAAGT 429

QY 424 GATGTTTCATCGCTAGCGGTTTTGAGACGCTGTGTTGAAGGCTGGGTGCTTATGCTGT 483

Db 430 TATGCCCGTATCTGATTTGTGTCCGACACAGAACTGGGACGCGTGGTGTCTATGCTGC 489

QY 484 TGGGAGCACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAAGATTCAT 543

Db 490 TGGGAGCATTTGTGCGCCCTTGAGCAAGTACGCGCTGTACTCCAGCATGAAGCCATTCC 549

QY 544 TGTCGGCTTGCGCGAGCTTTAGCCTTTATCTTAATGCGCGGAAAGCCCTGGGGCTGAT 603

Db 550 ATTGCTGCGCTGGCCGTCGTTTCGCTATACAGCGAACAGGCCCAAGCCCTCAGTGCCAA 609

QY 604 GTCAATGTAGCGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTTCTAGTACGCTGC 663

Db 610 GTGAACATGGCTGCTCGCAATCTATTGCGTTGAAGGCCAGTGCTTTACCATCGCGCC 669

QY 664 TGTGCGCTCGTTTCAACAATTCATGATGATATGCTTTGTACAGATGACGAAAAGCATGCG 723

Db 670 AGCAGTGTGTCAACCCAAAGACGCTAGACATGCTGGAAGTGGGTGAACACAAAGCCCC 729

QY 724 TTGCTTCTGCTGTGTGTGACACTCACGTATCATAGGGCTGATGTTGTGACTGTGC 783

Db 730 TTGCTGAAGAAGTGGCGGCGGCAATTCATGATTTTGGCGCCGACGACACACTGGCT 789

QY 784 GCGCCTCTTGCCGAAAATGAAGAGGATTTCTACGCAAACTTGATCCTGAGTACGC 843

Db 790 CCTACCTGCTCACGATGCCAGGGGCTTGATCATTTGCCGATCTGAATATGAGAGATTT 849

QY 844 ATCCTTGCTAAATGCGCGGACACCTGCTGTGCATTAATTCGCTCCCGACATTACTCGC 903

Db 850 GCCTTCGCCAAAGCATCAATGACCCCGTAGGCCACTATTTCCAAACCCGAGGCCACCCGT 909

QY 904 TTGCTAATAGA 914

Db 910 CTGGTGTCTGA 920

```

RESULT 2
US-10-919-182-15
; Sequence 15, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CI2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
;
; LENGTH: 1110
;

```

```

; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1110)
US-10-919-182-15

```

Query Match	21.7%;	Score 219.6;	DB 9;	Length 1110;
Best Local Similarity	52.1%;	Pred. No. 1.1e-60;		
Matches 489;	Conservative 0;	Mismatches 449;	Indels 0;	Gaps 0;

QY		22	GCCCTGCGTGAAGCCGCCCGATCTACATGGATTGGAGGCGA	C	CGGTGGACA	AAAACCA	T	81
Db		31	GCAACCCTTCAGGCAGAGCCGGTATGGCTCGACGCGAGA	C	CCA	CGATCGACA	AAGTGCATC	90
QY		82	GAGTTGATGAAGAAGCAGCACGTAATATGCTCGTCTGATCG	C	CTTCCG	AAACTTGG		141
Db		91	GGCATCATCGAAGAGCTGCCAAAAAGGCGCGAGTCTGATCG	CTTCCCG	AAATATTC			150
QY		142	ATTCCAGGCTACCCATGGTTTTCTTGCGTTGACTCA	CCAGCATGGG	CAATGCA	ATTGTA		201
Db		151	ATTCCGGGCTACCCCTATTGGCGGTGCTCGCGCAGTGA	AGTACAGCCTA	AGCTTTACT			210
QY		202	CGCCAATACCATGAGAACTCATTTGGAGTTGGATGG	CCCTCAAGCTA	AGCGCATTTCA	GAT		261
Db		211	TCACGCTATCACGAGAAATTCGTTGGAGCTAGGTGA	CGACCCGTATG	CGCTCCAGCTG			270
QY		262	GCAGCCAAAGCGGTGGGAATCATGGTCA	CCCTGGGGATGA	TGAACGGGT	CGSTGGCAC		321
Db		271	GCCGCGCGCCGCAACAAAATCGCACTCGTCATGG	CTATTCCG	AGCGGAAAG	CCGGATCG		330
QY		322	CTTTACATCAGTCAGTGGTTCATAGCGATAATG	TGA	CACCATTTGGGG	CCCCGCGCAAG		381
Db		331	CGCTATCTGAGCCAGGTGTTTCATCGACGAGCGTGG	CGCAGATCGTTG	CCAATCGGCGCAAG			390
QY		382	TTGAAACCTACTTTTGTGAAACGTACTTTGTT	CGCGCAAGGGAGT	GTTCA	TGCTAGCG		441
Db		391	CTGAAGCCCACACACAGTTGAGCGTAGATCTA	CGCGCAAGGCAACG	GAATTTCTC			450
QY		442	GTTTTGAGACGTCTGTTGGAAGCGTGGTGGCTTA	TGCTGTGGAG	CACCTTCA	ACCG		501
Db		451	ACGCACGACTTCGCGTTGCGACGCGTCG	TGATGAACTG	CTGGAACATGTTCA	ACCG		510
QY		502	CTAACAAATACGCTTGTATGACAAATGAAGA	GAATTCATTGTG	CGCGCTTG	CGCGAGC		561
Db		511	CTCAGCAAGTTCATGATGTATACAGCCTCG	GTGAGCAGGTCCA	CTTGCA	TGTCGCGCGG		570
QY		562	TTTAGCCTTATCCTAATGCGCGCAAAAGCC	CTGGGCGCTGATG	TCATGTAG	CGGCTCT		621
Db		571	ATGTCCCTCTTCAGCCGGATGTTTCCA	ACTGAGCATCGAA	GCCAACGCGA	CGGTCA		630
QY		622	CGAATCTATGCCGTTGAAGGGCAATGCTTC	GTA	CTAGCGTGTG	CGCTCGTTCA	CAA	681
Db		631	CGCTCGTACGCAATCGAAGGCCAAACCTT	TGTGCTTGCTGA	CGCAGGTG	ATCGGACCT		690
QY		682	TCCATGATCGATATGCTTTGTACAGATGACG	AAAGCATGCGT	GCTTCTG	GCTGTGT		741
Db		691	AGCGGATCGAAACGTTCTGCTCAACGAC	GAACAGCGGCA	CTGTTGCCG	MAAGATG		750
QY		742	GGACACTCACGTATCATAAGGCGCTGATGGT	GTGTA	CTTGCGC	CTCTTGCCG	AAAT	801
Db		751	GCGTGGGCGCGCATTTACGGCCCGATGGA	AGCGAGCTTG	CGAAGCCT	CTGCGG	AAAGAT	810
QY		802	GAAGAGGGTATTCTACGCAAACTTGATC	CTGAGTA	CGCATCCTTG	CTAA	AATGGCG	861
Db		811	GCTGAGGGGATCTTGACGACAGATCGAT	CTGAGCAGAT	TTCTGCTG	CGGAAG	CTGGA	870
QY		862	GCAGACCCTGCTGTCAATATCCCGTCCG	ACATTA	CTGCTT	GCTAATAG	ATCGCAGC	921
Db		871	GCGGATCCGTCGGGCACTATTGCGCGC	CTGACGTGCTG	TGCGTCCAG	TTGACCC	CGCGC	930

QY 922 CCTAAATTACCGTAGTTGAATGAAGTGATCTTCG 959  
Db 931 AATCATACGCCAGTTCATCGCATCGCATGTGACGCTCG 968

RESULT 3  
US-10-919-182-17

; Sequence 17, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 17  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-17

Query Match 21.7%; Score 219.6; DB 9; Length 1110;  
Best Local Similarity 52.1%; Pred. No. 1.1e-60;  
Matches 489; Conservative 0; Mismatches 449; Indels 0; Gaps 0;

QY 22 GCCTGCGTGCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGACAAAACCAATT 81  
Db 31 GCAACCGTTCAGGACAGCCGGTATGGCTCGACGACAGCGAATGATGACAGTGCATC 90  
QY 82 GAGTGTATGAAGAAGCAGCAGCTAATATGCTCTGATGATGCTTCCGGAACCTTGG 141  
Db 91 GGCATCATCGAAGAAGCTGCCAAAAGGCGCGAGTCTGATCGCTTCCGGAATATTC 150  
QY 142 ATTCCAGGCTTACCCATGTTCTTTGGCTTGACTCACCAAGCATGGGCAATGCAATTGTA 201  
Db 151 ATTCCGGGCTACCCCTATTGGCGGTGGCTCGGACGTGAAGTACAGCCTAAGCTTACT 210  
QY 202 CGCCATATACCATGAGACTCATTTGAGAGTTGATGGCCCTCAAGCTAAGCGCATTTCAAT 261  
Db 211 TCACGCTATCACGAGAATTCGTTGAGAGCTAGGTGACGACCGCTATGCGCTCCAGCTG 270  
QY 262 GCAGCCCAAGCGGTTGGGAATCATGTGTCACCCCTGGGGATGATGAACGGGTGGTGACCC 321  
Db 271 GCCGCGCGCCGCAACAATCGCACTGTCATGCGCTATTTCGGAAGCGGAAAGCCGGATCG 330  
QY 322 CTTTACATCAGTCAGTGGTTTCATAGGCGATAATGTGACACCATTTGGGCCCCGGCGAAG 381  
Db 331 CGCTATCTGAGCCAGGTGTTTCATGACGAGCGTGGCGAGATCGTTGCCAATCGCGCAAG 390  
QY 382 TTGAAACCTACTTTTGTGAAGTAATCTTTGTTCCGCGAAGGGAGTGTTCATCGCTACCG 441  
Db 391 CTGAAGCCACACACAGTTGAGGTACGATCTACGCGAAGGCAACGGAACCGATTCTC 450  
QY 442 GTTTTCGAGAGCTGTGTGAAGGCTGGGTGCTTATGCTGTGGAGACACCTTCAACCG 501  
Db 451 ACGCAGCACTTCGCGTTCGACCGGTGCGGTGATTTGAATGCTGGGAACATCTACACCG 510  
QY 502 CTAAACAATAATACGCTTTGTATGACAAATAAGAGAGATTATTTGCGGCTTGCGCCAGC 561  
Db 511 CTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAAGGTCCACGTTGCATCGTGGCCGCG 570  
QY 562 TTTAGCCTTTATCTAATGCGGGAAGCCCTGGGGCTGATGTCAATGTAGCGGCTCT 621

Db 571 ATGTCCTCTTACGCCGAGTGTTCCTCAACTGAGCATCGAAGCCAAAGCGACGGTCACC 630  
QY 622 CGAATCTATGCCGTTGAAGGGCAATGCTTCTGACTAGCCTGCTGCGCTGTTTCAAA 681  
Db 631 CGCTGTACGCAATCGAAGGCCAAACCTTTGCTTGTCTGACGACGATGATCGACCT 690  
QY 682 TCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTGGCTGTGGT 741  
Db 691 ACGCGCATCGAAACGTTCTGCTCAACGACGAAAGCGGCACTGTGCGCGAAGATGT 750  
QY 742 GGACACTACGATATCATAGGGCCCTGATGTTGTTGCTTGGCTTGGCCGAAAT 801  
Db 751 GGCTGGGCGGCAATTCAGGCGCCGATGGAAGCGACTTGGGAAGCCTTGGCGGAAGAT 810  
QY 802 GAAGAGGTATTTCTACGCAAACTTGATCTTGAGTACGCATCTTGCTAATAATGCGG 861  
Db 811 GCTGAGGGATCTTGTACGACAGATGATCTGAGCAGATTTCTGTCGGCGAAGCTTGA 870  
QY 862 GCAGACCCCTGCTGTCATTTATTCCTCCGTCGCCGACATTAATCTGCTAATAGATCGAC 921  
Db 871 GCCGATCCGTCGGGCACTATTCGCGCCTGACGTCGTCTGCTCCAGTTGACCCGCGC 930  
QY 922 CCTAAATTACCGTAGTTGAATGAAGTGATCTTCG 959  
Db 931 AATCATACGCCAGTTCATCGCATCGCATGTGACGCTCG 968

RESULT 4  
US-10-919-182-3

; Sequence 3, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Acidovorax facilis 72W  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-3

Query Match 21.5%; Score 218; DB 9; Length 1110;  
Best Local Similarity 52.0%; Pred. No. 3.7e-60;  
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

QY 22 GCCTGCGTGCAAGCCGCCGATCTACATGATTTGAAGGCGACGGTGACAAAACCAATT 81  
Db 31 GCAACCGTTCAGGACAGCCGGTATGCTTCGACGACGACGCAACGATGACAAATGATC 90  
QY 82 GAGTGTATGAAGAAGCAGCAGTAATATGCTGCTGATGCTTCCGGAACCTTGG 141  
Db 91 GGCATCATCGAAGAAGCTGCCAAAAGGCGCGAGTCTGATCGCTTCCCGGAATATTC 150  
QY 142 ATTCCAGGCTACCCATGTTCTTTGCTTGACTCACCAAGCATGGCAATGCAATTGTA 201  
Db 151 ATTCCGGGCTACCCCTATTGGGCGGTGCTCGGCAAGTGAAGTACAGCCTAAGCTTACT 210  
QY 202 CGCCATATACATGAGAACTCATTTGAGTTGATGCGCCCTCAAGCTAAGCGATTTCAGAT 261  
Db 211 TCACGCTATCACGAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGTCCCTCCAGCTG 270  
QY 262 GCAGCCAAGCGGTTGGGAATCATGTACACCTTGGGGATGATGAACGGGTGGTGACCC 321

```

Db      271  GCCGCGCGCCGCAACAAATCGCACTCGTCATGGGCTATTGAGCGGGAGCCGGATCG 330
QY      322  CTTTACATCAGTGAAGTGTTCATAGCGCATATAGTGACACCAATTGGGCGCGGAAAG 381
Db      331  CGCTATCTGAGCCAGGTGTTCATGACGAGCGTGGCGAGATCGTCCCAATCGCGCAAG 390
QY      382  TTGAAACCTACTTTTGTGTAACGACTTTGTTCCGCCGAAGGGAGTGTTCATCGTAGCG 441
Db      391  CTGAAGCCCAACACACGCTTGAGCGTACGATCTACGGCGAAGGCAACGAAATTCCTC 450
QY      442  GTTTTCGAGACGCTGTGTTGGAAGGCTGGGTGCTTATGCTGTGGAAGACCTTCAACCG 501
Db      451  ACGCAGCACTTCGCGTTCGGAACGCGTCGCTGGAATTGAATGCTGGGAACATTTCCAA 510
QY      502  CTAACAAATATACGCTTTGTATGCACAAATGAAGATTCATTGCGGCTTGGCCGAGC 561
Db      511  CTCAGCAAGTTCAATGATGATACAGCCTCGGTGAGCAGGTCCACGTTGCATCGTGGCC 570
QY      562  TTTAGCCTTTATCCTTAATGCGGCGAAAGCCCTGGGCGCTGATGTCATGTAGCGGCT 621
Db      571  ATGTCCCTCTTCAGCCGATGTTTTCCAACTGAGCATCGAAGCCACGCGAGGTCAACC 630
QY      622  CGAATCTATGCGGTTGAAGGCAATGCTTCTGACTAGCGCTGCTGCGCTCGTTTCAACA 681
Db      631  CGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTCTGACGAGGTGATCGGAACC 690
QY      682  TCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGCTTCTGCTGCTGGT 741
Db      691  AGCGCGATCGAAACGTTCTGCTCAACGACGAAACGCGCACTGTGCGCGAAGATGT 750
QY      742  GGACACTACGATATCATAGGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 801
Db      751  GGCTGGCGCGCATTTACGCGCCCGGATGGAAGCGAGCTTGCGAAGCCCTTGCGCGAAG 810
QY      802  GAAGAGGATTTCTCTACGCAACCTTGATCTCTGAGTACGCAATCTTGTCTAAATGGCG 861
Db      811  GCTGAGGGGATCTTGTACGACAGATGATCTGAGCAGATTCCTGCGGAAGGCTGGA 870
QY      862  GCAGACCTGCTGTCATTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921
Db      871  GCCGATCCGGTGGGCACTATTGCGCGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
QY      922  CCTAATTAACCGGTAGTTGAATTTGAAGGTGATCTTCG 959
Db      931  AATCATACGCCAGTTCATCGCATCGGCATTTGACGGTGC 968

```

```

RESULT 5
US-10-919-182-5
; Sequence 5, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: DI Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.2
; SEQ ID NO 5
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B2 and H9
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-5

```

```

Query Match      21.5%; Score 218; DB 9; Length 1110;
Best Local Similarity 52.0%; Pred. No. 3.7e-60;
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

QY      22  GCCTGCTGCAAGCCGCCGATCTACATGATTGAGGCGGAGGTGACAAACCAT 81
Db      31  GCAACCGTTACAGGAGAGCCGATGCTGACGACGACGCAACGATCGAACAGTGCATC 90
QY      82  GAGTGTATGGAAGACAGCAGTAATATGCTGCTGATGCTGCTGCTGCTGCTGCTGCTG 141
Db      91  GGCATCATCGAAGAGCTGCCCAAAAGGCGCGAGTCTGATGCTTTCCCGGAATATTC 150
QY      142  ATTCCAGGCTACCCATGGTTTCTTGGCTTGACTCAACGATGGGCAATGCAATTGTA 201
Db      151  ATTCCGGGCTACCCCTATTGGCGCTGCTGCGGACGTGAAGTACAGCTTAAGCTTACT 210
QY      202  CGCCAATACCATGAGACTCATTTGAGTTGATGCGCCCTCAAGCTTAAGCGCATTTAGAT 261
Db      211  TCACGCTATCACGAAATTCGTTGAGCTAGGTGACGACCGTATGCTGCTCCAGCTG 270
QY      262  GCAGCCAAGCGGTTGGAAATCATGCTACCCCTGGGAGTGAAGTGAACGGGTGCGCAC 321
Db      271  GCCCGCGCCGCAACAAATCGCACTCGTCATGGCTAATTCGAGCGGGAAGCCGGATCG 330
QY      322  CTTTACATCAGTCACTGCTTCAATAGGCGTAATAGTGAACCATTTGGGCGCGGAAAG 381
Db      331  CGCTATCTGAGCCAGGTGTTCAACGACGAGCGTGGGAGATCGTTGCCAATGCGCAAG 390
QY      382  TTGAACCTACTTTTGTGTAAGCTATTGTTGCGGCAAGGGAGTGTTCATGCTAGCG 441
Db      391  CTGAAGCCCAACACAGTTGAGCTGATCTACGCGGAAGCAACGAAACGATTTCTC 450
QY      442  GTTTTCGAGACGCTGTTGGAAGGCTGGGTGCTTATGCTGTTGGAGCACCTTCAACG 501
Db      451  ACGCAGCACTTCGCGTTCGACGCGCTGCTGATTTGAATGCTGGGAACATTTCCAACG 510
QY      502  CTAACAAATATACGCTTTGTATGCAAAATGAAGATTCATTGCGGCTTGGCGGAGC 561
Db      511  CTCAGCAAGTTCAATGATGTAACAGCTTCGTTGAGCAGGTCCAGTTGCATGCTGGCC 570
QY      562  TTTAGCTTTATCTTAATGCGGCGAAAGCCCTGGGCGCTGATGTCATGTAAGCGCTCT 621
Db      571  ATGTCCCTCTTCAGCCGAGTGTTCCTCAACTGAGCATCGAAGCCAACGCGAGTGC 630
QY      622  CGAATCTATGCGGTTGAAGGCAATGCTTCTGTAAGCGCTGCTGCTGCTGCTGCTGCT 681
Db      631  CGCTGTAACGCAATCGAAGGCCAAACCTTTGTGCTTGTCTGACGCGATGATCGACCT 690
QY      682  TCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGCTTCTGCTGCTGCTGCT 741
Db      691  AGCGGATCGAAACGTTCTGCTCAACGACGAAACAGCGGCACTGTTCCGCAAGATGT 750
QY      742  GGACACTACGATATCATAGGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 801
Db      751  GGCTGGCGCGCATTTACGGCCCGGATGGAAGCGAGCTTGCGAAGCCTTCTGCGGAAGAT 810
QY      802  GAAGAGGATTTCTCTACGCAACCTTGATCTCTGAGTACGCAATCCTTGTCTAAATGGCG 861
Db      811  GCTGAGGGGATCTTGTACGAGATGATGATCTGAGCAGATTCCTGCTGGCGAAGCTGGA 870
QY      862  GCAGACCTGCTGTCATTATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 921
Db      871  GCCGATCCGGTGGGCACTATTGCGCGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
QY      922  CCTAATTAACCGGTAGTTGAATTTGAAGGTGATCTTCG 959
Db      931  AATCATACGCCAGTTCATCGCATCGGCATTTGACGGTGC 968

```

```

RESULT 6
US-10-919-182-11
; Sequence 11, Application US/10919182
; Publication No. US20060035352A1

```



```
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
; US-10-919-182-11
```

```
Query Match          21.5%; Score 218; DB 9; Length 1110;
Best Local Similarity 52.0%; Pred. No. 3.7e-60;
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;
```

```
QY 22 GCCTGCGTGAAGCCGCGCCGATCTACATGGATTGGAGCGCGATGGACAAACCATT 81
   |||||
Db 31 GCAACCGTTGAGGACAGAGCCGGTATGGCTCGACGACGCGCAACGATCGACAAAGTCATC 90
QY 82 GAGTTGATGGAAGAAGCAGCAGCTAATATATGCTGCTGATCGCTTCCGAAACTTGG 141
   |||||
Db 91 GGCATCATCGAAGAAGCTGCCAAAGGGCGCGAGTCTGATCGCTTCCCGAAGTATTC 150
QY 142 ATTCAGGCTACCCATGTTCTTGGCTTGACTCACCAGCATGGGCAATGCAATTTGTA 201
   |||||
Db 151 ATTCGGGGCTACCCCTATGGCGCTGCTCGGCGACGTGAAGTACAGCCTAAGCTTTACT 210
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCATTTCAAGT 261
   |||||
Db 211 TCAGGCTATCACGAGAATTCGTTGAGCTAGGTGACGACCGTATGCGTCCAGCTG 270
QY 262 GCAGCCAAGCGGTTGGAATCATGGTCAACCTGGGGATGATGAACGGGTCCGTGGCAC 321
   |||||
Db 271 GCCGCGCGCCGCAACAAATCGCACTGTCATGGGCTATTCGAGCGGGAAGCCGATCG 330
QY 322 CTTTACATCAGTCAGTGGTTCATAGGCGATATATGTTGACACCATTTGGGCGCGCAAG 381
   |||||
Db 331 CGCTATCTGAGCCAGGTGTTCACTGACGAGCGTGGCGAGATCTTGCCAATCGCGCAAG 390
QY 382 TTGAAACCTACTTTTGTGAACGTACTTTGTTGCGGGAAGGAGATGTTTCATCGCTAGCG 441
   |||||
Db 391 CTGAAGCCCAACACACGTTGAGCGTACGATCTACGCGAAGGCAACGGAACGATTTCTC 450
QY 442 GTTTGAGAGCGTCTGTTGAAGGCTGGGTGCTTATGCTGTTGGAGACACCTTCAACG 501
   |||||
Db 451 ACGCAGCACTTCGCGTTCGACGCGTCCGTGATTTGAATGCTGGGAACATTTCCAACG 510
QY 502 CTAAACAATAAGCTTTGTATGCAACAATAAGAGATTCATTGCGGCTTGGCCGAGC 561
   |||||
Db 511 CTGACGAAGTTCATGATGATGACGCTCGGTGAGCAGGTCACGTTGCATCGTGGCCGCG 570
QY 562 TTAGCCTTATCTAATGCGGCGAAAGCCCTGGGGCTGATGTCAATGTAGCGGCTCT 621
   |||||
Db 571 ATGTCCCTCTTCAGCCGAGTGTTCCTCAACTGAGCATGCAAGCCAACGCGAGGTCGC 630
QY 622 CGAATCTATGCCGTTGAAGGCAATGCTGTAAGCTGCTGCTGCTGCTGCTTTCACAA 681
   |||||
Db 631 CGCTCGTACGCAATCGAAGGCCAAACCTTTGTGCTTGTGACGAGGTGATCGGAGCT 690
QY 682 TCCATGATCGATATGCTTTGTAAGATGACGAAAGCATGCGTGTGCTTGTGCTGCTGCT 741
   |||||
Db 691 AGCGGATCGAAACGTTCTGCTCAACGACGAAACAGCGCGCACTGTTGCCGAAGATGT 750
```

```
QY 742 GGAACCTACGATCATAGGCGCTGATGTGTGACTTGGTCCGCGCTCTTCCGAAAT 801
   |||||
Db 751 GGCTGGCGCGCGCATTTACGCGCCGAGATGGAAGCGAGCTTGGCAAGCCTCTGGCGAAGAT 810
QY 802 GAAGAGGATTTCTTACGCAAACTTGATCCTCGAGTACGCATCTTGCTAAATGCGC 861
   |||||
Db 811 GCTGAGGGGATCTTGTACGAGAGATCATCTGAGCAGATTTCTGCTGGCGAAGCTGA 870
QY 862 GCAGACCTGCTGTCATTAATCCCGTCCGACATTAATCGCTTGTATAGATCGCAGC 921
   |||||
Db 871 GCCGATCCGGTCCGGCACTAATTCGGCGCTGACGCTGCTGCTGCTGCTGCTGCTGCTG 930
QY 922 CCTAATTACCGGTAGTTGAATGAAGTGATCTCG 959
   |||||
Db 931 AATCATACGCCAGTTTCATCGCATCGCATTTGACGCTCG 968
```

## RESULT 7

```
US-10-919-182-7
; Sequence 7, Application US/10919182
; Publication No. US2006003532A1
```

## ; GENERAL INFORMATION:

```
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
```

```
; APPLICANT: Payne, Mark
```

```
; APPLICANT: O'Keefe, Daniel
```

```
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
```

```
; FILE REFERENCE: CL2584 US NA
```

```
; CURRENT APPLICATION NUMBER: US/10/919,182
```

```
; NUMBER OF SEQ ID NOS: 18
```

```
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 7
```

```
; LENGTH: 1110
```

```
; TYPE: DNA
```

```
; ORGANISM: Artificial sequence
```

```
; OTHER INFORMATION: Mutant nitrilase B4
```

```
; FEATURE:
```

```
; NAME/KEY: CDS
```

```
; LOCATION: (1)..(1110)
```

```
; US-10-919-182-7
```

```
Query Match          21.3%; Score 216.4; DB 9; Length 1110;
Best Local Similarity 51.9%; Pred. No. 1.2e-59;
Matches 487; Conservative 0; Mismatches 451; Indels 0; Gaps 0;
```

```
QY 22 GCCTGCGTGAAGCCGCGCCGATCTACATGGATTGGAGCGCGATGGACAAACCATT 81
   |||||
Db 31 GCAACCGTTGAGGACAGAGCCGGTATGGCTCGACGACGCAACGATCGACAAAGTCATC 90
QY 82 GAGTTGATGGAAGAAGCAGCAGTAATAATGCTGCTGATCGCCTTCCGAAACTTGG 141
   |||||
Db 91 GGCATCATCGAAGAAGCTGCCAAAGGCGCGAGTCTGATCGCTTCCCGAAGTATTC 150
QY 142 ATTCAGGCTACCCATGCTTCTTGGCTTGACTCACCAGCATGGGCAATGCAATTTGTA 201
   |||||
Db 151 ATTCGGGGCTACCCCTATTTGGGCGTGGCTCGGCGACGTGAAGTGCAGCCTAAGCTTTACT 210
QY 202 CGCCAATACCATGAGAACTCATTTGAGTTGGATGGCCCTCAAGCTAAGCGCATTTCAAGT 261
   |||||
Db 211 TCAGGCTATCACGAGAATTCGTTGAGCTAGGTGACGACCGTATGCTGCTGCTCAGCTG 270
QY 262 GCAGCCAAGCGGTTGGAATCATGTCACCTCGGCGAGTGAAGTGAACGGGTCCGTGGCAC 321
   |||||
Db 271 GCCGCGCGCCGCAACAAATCGCACTGTCATGAGCTAATTCGAGCGGGAAGCGGATCG 330
QY 322 CTTTACATCAGTCAGTGGTTCATAGGCGATATATGTTGACACCATTTGGGCGCGCAAG 381
   |||||
Db 331 CGCTATCTGAGCCAGGTGTTTCATGACGAGCGTGGCGAGATCGTTGCCAATCGCGCAAG 390
QY 382 TTGAAACCTACTTTTGTGAACGTACTTTGTTGCGGCGAAGGAGTGTTCATCGCTAGCG 441
```



```
Db      391 CTGAAGCCCAACACGTTGAGCGTACGATCTACGGCAAGGCAACGGAACCGATTCTC 450
QY      442 GTTTTCGAGACGCTCTGTGGAAGCGTGGGCTTATGCTGTGGAGCACTTCAACCG 501
Db      451 ACGCAGCACTTCGGGTCGGAACGCGTCGGTGAATTGAACCTGCTGGGAACATTTCAACCG 510
QY      502 CTACAAATAACGCTTTGTATGCACAAATGAAGATTGATGTCGGCTGGCCGAGC 561
Db      511 CTCAGCAAGATCATGATGTACAGCCTCGGTGAGCAGGTCCACGTTGCATCGTGGCCGCG 570
QY      562 TTTAGCCTTTATCCTAATGCGGCGAAAGCCCTGGGCGCTGATGTCATGTAGCGGCTCT 621
Db      571 ATGTCCCTCTTCAGCCGGATGTTTCCAACTGAGCATCGAAGCCAACGCGACGGTCATC 630
QY      622 CGAATCTATGCCGTTGAAGGCAATGCTTCGTAAGCGTCGTGTCGCTGCTTTCACAA 681
Db      631 CGCTCGTACGCAATCGAAGGCCAACCTTGTGCTTGTCTCGACGCAAGTGAATCGAACCT 690
QY      682 TCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGCGCTTCTGGCTGGTGT 741
Db      691 AGCGCGATCGAAACGTTCTGCGCTCAACGACGAAACGCGCACTGTGCGCGAAGATGT 750
QY      742 GGACACTCAGCTATCATAGGCGCTGATGCTGTGCTGCTGCGCTCTTGGCGAAAT 801
Db      751 GCGTGGCGCGCATTTTACGCGCCGATGGAAGCGAGCTTGGCAAGCCTCTGGCGAAGAT 810
QY      802 GAAGAGGGTATTTCTTACGCAAACTTGATCTGAGTACGATCCTTGTAATAATGGCG 861
Db      811 GCTGAGGGGATCTTGTACGACAGATCGATCTGAGCAGATTTCTGCGCGAAGGTGGA 870
QY      862 GCAGACCTCTGCTGTCTATTATTCGCTCCGACATTACTGCTTGTCTAATAGATCGCAGC 921
Db      871 GCCGATCCGCTGGGCACTATTTCGGCGCTGACGTCGTGCTGCTCAGTTGCACCCGCGC 930
QY      922 CCTAAATTACCGGTAGTTGAATGAAGGTGATCTTCG 959
Db      931 AATCATACGCCAGTTTCATCGCATCGCATTTGACGGTCCG 968
```

```
RESULT 8
US-10-919-182-13
; Sequence 13, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: DI Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-13
```

```
Query Match      21.3%; Score 216.4; DB 9; Length 1110;
Best Local Similarity 51.9%; Pred. No. 1.2e-59;
Matches 487; Conservative 0; Mismatches 451; Indels 0; Gaps 0;
```

```
QY      22 GCCTGCGTGCAAGCCCGCCGATCTACATGATTTGGAGCGACGCTGACAAACCAT 81
; Db      31 GCAACCGTTCAGGCAAGCCCGGTATGGCTGACGCAAGCAAGATCGACAAGTCGATC 90
```

```
QY      82 GAGTTGATGGAAGACGACGTAATATGCTCTGATCGCCTTTCCGGAACCTGG 141
Db      91 GGCATCATCGAAGAAGCTGCCCAAAAGGGCGCGAGTCTGATCGCTTTCCCGGAAGATTC 150
QY      142 ATTCAGGCTACCCATGGTTCTTTGGCTTGACTCAACGATGGGCAATGCATTTGTA 201
Db      151 ATTCGGGGCTACCCCTATTTGGGCGTGGCTCGCGCAGCTGAAGTACAGCCTTAAGCTTTACT 210
QY      202 CGCAATACCATGAGAATCTGAGTTGAGTTGATGGCCCTCAAGCTAAGCGCATTTCAAGT 261
Db      211 TCAGCTATCACGAAATTCGTTGAGCTAGTGACGACCGTATGCGTCCCTCCAGCTG 270
QY      262 GCAGCAAGCGGTTGGAAATCATGTCTACCCCTGGGAGTGAACGGGTGGTGCAAC 321
Db      271 GCCCGCGCCGCAACAAATCGCACTCGTCATGGGCTAATTCGAGCGGGAAGCCGATCG 330
QY      322 CTTTACATCAGTCAGTGGTTCAATAGCGGATAATGATGACCATTTGGGCGCCGCGAAAG 381
Db      331 CGCTATCTGAGCCAGGTGTTCAATCGACGAGCGTGCGAGATCGTTGCCAATCGCGCGAAG 390
QY      382 TTGAACCTACTTTTGTGTAACGTACTTTGTTCCGCGAAGGAGTGTTCATCGCTAGCG 441
Db      391 CTGAAGCCCAACACACGTTGAGCGTACATCTACGCGAAGGCAACGAAACGATTTCTC 450
QY      442 GTTTTCGAGACGCTCTGTGGAAGGCTGGGTGGCTTATGCTGTTGGAGACACCTTCAACCG 501
Db      451 ACGCAGCACTTCGCGCTTCGGAACGCGCTGGTGGATTGAACGCTGCGGAACATAAACCAACCG 510
QY      502 CTACAAATAACGCTTTGTATGCACAAATGAAGAGATTCAATGCGGCTTGCGCGAGC 561
Db      511 CTCAGCAAGTTCATGATGTACAGCCTCGGTGAGCAGAGTCCACGTTGCATCGTGCGCGCG 570
QY      562 TTTAGCCTTTATCTTATATGCGGCGAAAGCCCTGGGCGCTGATGTCATATGAGCGGCTCT 621
Db      571 ATGTCCCTCTTCAACCGGATGTTTTCGAACCTGAGCATCGAAGCCAACCGGAGTCAAC 630
QY      622 CGAATCTATGCCGTTGAAGGCAATGCTTGTACTAGCGTCGTGCGCTCGTTTCAACA 681
Db      631 CGCTGATACGCAATCGAAGGCCAAACCTTGTGCTTGTCTGACGCGAGTGAATCGGACCT 690
QY      682 TCCATGATGATATGCTTTGTACAGATGACGAAAGCATGCGTTCCTTCTGGCTGTGT 741
Db      691 AGCGCGATCGAAACGTTCTGCTCAACGACGAAACAGCGCACTGTGCGCGAAGATGT 750
QY      742 GGACACTCAGCTATCATAGGCGCTGATGCTGTGTAACCTTGTGCGCGCTTGGCGAAAT 801
Db      751 GCGTGGCGCGCATTTTACGCGCCGATGGAAGCGAGCTTGCGAAGCCTTGGCGGAAGAT 810
QY      802 GAAGAGGGTATTTCTTACGCAAACTTGATCTGAGTACGATCCTTGCTAAATGGCG 861
Db      811 GCTGAGGGGATCTTGTACGACAGATCGATCTGAGCAGATTTCTGCGCGAAGGTGGA 870
QY      862 GCAGACCTCTGCTGTCTATTATTCGCTCCGACATTACTCGCTTGTAAATGATCGCAGC 921
Db      871 GCCGATCCGCTGGGCACTATTTCGGCGCTGACGTCGTGCTCAAGTTGCACCCGCGC 930
QY      922 CCTAAATTACCGGTAGTTGAATGAAGGTGATCTTCG 959
Db      931 AATCATACGCCAGTTTCATCGCATCGGCATTTGACGGTCCG 968
```

```
RESULT 9
US-11-096-568A-6953
; Sequence 6953, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
```

```
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6953
; LENGTH: 1415
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1415)
; OTHER INFORMATION: Ceres Seq. ID no. 15169316
US-11-096-568A-6953
```

```
Query Match      8.2%; Score 83; DB 11; Length 1415;
Best Local Similarity 50.1%; Pred. No. 4.5e-16;
Matches 271; Conservative 0; Mismatches 255; Indels 15; Gaps 2;
```

```
QY 23 CCTGCGTGCAAGCCGCCCATCTACATGATTTGGAGGCGACGGTGACAAACCAATTG 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 CCGTGTGTTCAAGCCTCCACCACCTTCTACGACACACACCAACCCTTAGATAAGGCTGAGA 244
QY 83 AGTTGATGGAAGAAGCAGCAGCACTAATATGCTCGTCTGATCGCCTTCCGAAACTTGG 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 GGTGTTGGCTGAAGCTACTAGCTATGGGTCCACGCTTGTTGTTCCAGAAGCCTTTG 304
QY 143 TTCCAGGCTACCCCATGTTT---CTTGGCTTGACTCAGCAGCATGGGCAATGCAATT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 TGGGTGGCTACCCGCGTGGTTCACTTTTGGTCTTTCGATCGGAAATCGCACTGTTAAG 364
QY 199 GTA-----CGCCAATACCATGAGAACTCATGAGTTGAGTTGGATGGCCCTCAAGCT 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GTAGAGAGAGATTTCGCAAGTATCATTTCTGCAGCCATGATGTCCTGCTCGAAGTTG 424
QY 248 AGCGCATTTAGATGACAGCCAGCGGTTGGGAATCATGGTCAACCCTGGGATGATGAAC 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 ATAGATTGGCAGCAATGGCAGGGAAGTATAAGTACATTTAGTATGGGTGTGATAGAGA 484
QY 308 GGGTCGGTGGCAACCCTTTACATCAGTCAGTGTTCATAGGCGCATATGTTGACACCAATT 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 485 GGGACGGCTACACACTTATTTGACCCGTTCTGTTCTTTGATTTCTCAGGGTCACTTAAC 544
QY 368 GGGCCCCGGGAAAGTTGAACCTACTTTTGTGTAACGTACTTTGTCGGCGAAGGGATG 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 545 GAAAGCACAGGAAATCATGCAACTGCATTGGAGCGGGTATCTGGGATTTGGGATG 604
QY 428 GTTCATCGCTAGCGGTTTTCAGACGTCGTGTTGGAAGCTGGGCTTATGCTGTTGGG 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 605 GATCAACCATTCAGATGTTGAAGTCTCCCGTTGAAATAAGTGTGCTCCATTGTTGGG 664
QY 488 AGCACCTTCAACCGCTAACAATAACGCTTTGTATGCAAAATGAAGATTCATTGTG 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 AGAACAGATGCCACTATTAAGACAGCAATGTATGCCAAAGGTGTGAGATATATTGTG 724
QY 548 C 548
    |
Db 725 C 725
```

```
RESULT 10
US-11-096-568A-20685
; Sequence 20685, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20685
; LENGTH: 1706
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (1)..(1706)
; OTHER INFORMATION: Ceres Seq. ID no. 12386684
US-11-096-568A-20685
```

```
Query Match      8.0%; Score 81.6; DB 11; Length 1706;
Best Local Similarity 49.8%; Pred. No. 1.4e-15;
Matches 272; Conservative 0; Mismatches 259; Indels 15; Gaps 2;
```

```
QY 18 GGTCCCTGCGTGCAAGCCGCCCATCTACATGATTTGGAGGCGACGGTGACAAAC 77
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 GGTCAACCGTGTTCAGCGGTGTCGCTGTTCTAGACACCCCTGCACTCTGATAAAGC 688
QY 78 CATGAGTTGATGGAAGACGACGCACTAATATGCTCGTCTGATCGCCTTCCGGAAC 137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 GGAAGAATTGTAGCAGAAGCAGCTGGTATGTTCAAGTTGTTGTTCCGGAAGT 748
QY 138 TTGATTCAGGCTACCA-----TGATTCTTTGGCTTGACTCAC-CAGC 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 CTTTGTGTTGCTTACCCACATGATCTACTTGGAACTGTTGTGCGCAATCGCACTGC 808
QY 183 ATGGCAATGCAATTGTATCGCCAATACCATGAGAACTCATTTGAGTTGATGGCCCTCA 242
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 809 CAAGGAAAGAAAGACTTTCAGAAGTATCAGCATCTGCCATGATGTGCTGTCCAGA 868
QY 243 AGCTAAGCGCATTTAGATGACGCCAAGCGGTGGAAATCATGTGACCCCTGGGATGAG 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 869 AGTGTCCGCTTATCTGCAATTAGCTGAAATAATAGGTGTTTCTGTGATGGGCTGT 928
QY 303 TGAACGGGTGGGACCCCTTACATCAGTCAGTGTTCATAGGCGATATGTGACAC 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 929 TGAAGGGCAGATATACACTTTACACACGCTGCTCCTTGATCCACTGGGAAATA 988
QY 363 CATTTGGGCCCCGCAAGTTGAAACCTACTTTTGTGAACGTAATTGTTCCGCCAAG 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 989 CTTAGGGAAGCACCGCAAGGTCAATGCTTACTGCACTAGAACGTGATTTCTGGGGTTGG 1048
QY 423 GGATGTTATCGTACGCGGTTTTCAGACGTCGTGTTGGAAGCTGGGCTTATGCTG 482
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1049 AGATGATCTACATACTGTTCTACGATCTCCGATTGGAATAAGGTGCTCTCATTTG 1108
QY 483 TTGGAGCACCTTCAACCGCTAACAATAACGCTTTGTATGACAAATGAGAGATTCA 542
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1109 CTGGAAAAACAGAAATGCCACTTCTCAGACGGCATGTATGCCAAAGGTATGAGATATA 1168
QY 543 TTGTGC 548
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1169 TTGTGC 1174
```

```
RESULT 11
US-10-932-182A-166788
; Sequence 166788, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166788
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166788
```

```
Query Match      6.9%; Score 70.4; DB 9; Length 597;
```

[illegible]

```

RESULT 12
US-10-932-182A-166788
; Sequence 166788, Application US/10932182A
; Publication No. US20060046253A1
; GENERAL INFORMATION:
; APPLICANT: NAKAO, YOSHIHIRO
; APPLICANT: NAKAMURA, NORIHISA
; APPLICANT: KODAMA, YUKIKO
; APPLICANT: FUJIMURA, TOMOKO
; APPLICANT: ASHIKARI, TOSHIHIKO
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS
; FILE REFERENCE: 030685-043
; CURRENT APPLICATION NUMBER: US/10/932,182A
; CURRENT FILING DATE: 2004-09-02
; NUMBER OF SEQ ID NOS: 197023
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 166788
; LENGTH: 597
; TYPE: DNA
; ORGANISM: Saccharomyces pastorianus
US-10-932-182A-166788

```

	Query Match	6.9%;	Score 70.4;	DB 9;	Length 597;	
	Best Local Similarity	54.7%;	Pred. No.	3.9e-12;		
	Matches 140;	Conservative	0;	Mismatches 116;	Indels 0;	Gaps 0
QY	293 TGGGATGAGTGAACGGGTCCGTGCACACCCTTTACATCAGTCAGTGGTTTCATAGCGGATA					352
Dd	317 TGGGGTGATAGAGCGCGATGGAGCAACATTATTGTACCATGGTTTATATAGATCCTA					376
QY	353 ATGTGACACCATTTGGGCCCGCGAAAGTTGAAACTACTTTTGTTGAACGTACTTTGT					412
Dd	377 AAGATGGCTACGTTGGGAAGCATCGAAAAC TGATGCCGACAGCTGGCGAAAGACTGATAT					436
QY	413 TCGCGAAGGGGATGGTTTCATCGCTACGGGTTTTCGAGACGTCGTGGGAAGCGTGGGTG					472
Dd	437 GGGGTCAAGCCGATGGTTTCGACTCTGCTGTGCTGATACCGCTGCTGGGAAGATTGGCG					496
QY	473 GCCTATGCTCTTTGGGAGCACCTTCAACCGCTAACAAATACGCTTTGTATGCACAAATG					532
Dd	497 GTGCTATCTGCTGGGAGAACAATGATGCTCTTA CTGAGATACGCCCATGTATAAAAAAGGGG					556
QY	533 AAGAGATTCA TTGTGC 548					
Dd	557 TTGAGATCTGGTGTGC 572					

RESULT 13  
US-09-925-065A-628387

```

; Sequence 628387, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 628387
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-628387

```

Query Match	3.4%;	Score 34.6;	DB 6;	Length 548;
Best Local Similarity	56.6%;	Pred. No. 1.8;		
Matches	64;	Conservative	0;	Mismatches 49; Indels 0; Gaps 0
QY	71	ACAAACCATTTGAGTTGATGGAAGAAGCAGCAGCAGTAATAATGCTGCTGATGCGCTTTC	130	
Db	42	ACACAAAGATGAGTTCCTGGACATGAGTGACCTCAAAGATAACTCTGGGGACCTCTC	101	
QY	131	CGGAAACTTGATTTCCAGGCTACCCATGTTTCTTTGGCTTGACTCACCAGCA	183	
Db	102	TTTAGACCTGCAGCCTTTGGCTGGGCATGATGCTCATGCTGTAAACCCACGCA	154	

```

RESULT 14
US-09-925-065A-425677/c
; Sequence 425677, Application US/09925065A
; Publication No. US20040181048A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925, 065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243, 096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252, 147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250, 092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261, 766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289, 846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 425677
; LENGTH: 548
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-425677

```

Query Match	3.4%	Score 34;	DB 6;	Length 548;
Best Local Similarity	56.1%;	Pred. No. 2.8;		
Matches 64;	Conservative	0;	Mismatches 50;	Indels 0;

```
OY      473 GCTTATGCTGTGGAGCACCCTTCAACCGCTTAACAATAACGCTTTGTATGACAAATG 532
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      417 GATCATGCTACATGTTCTATCTTAACAGAAACAAATGAATTTTCTGGCTAAGCA 358
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      533 AAGAGATTCAATTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGA 586
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      357 AAATAATTCTTGGCTCTACTGGTCCCAATTTTGGCTATATCAAACTTAGTCCA 304
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 15

```
US-10-301-480-489364/C
; Sequence 489364, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; TITLE OF INVENTION: in the Human Genome
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489364
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-489364
```

```
Query Match      3.4%; Score 34; DB 10; Length 563;
Best Local Similarity 56.1%; Pred. No. 2.9;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```
OY      473 GCTTATGCTGTGGAGCACCCTTCAACCGCTTAACAATAACGCTTTGTATGACAAATG 532
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      417 GATCATGCTACATGTTCTATCTTAACAGAAACAAATGAATTTTCTGGCTAAGCA 358
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      533 AAGAGATTCAATTGCGGCTTGCGCGAGCTTTAGCCTTTATCCTAATGCGGCGA 586
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      357 AAATAATTCTTGGCTCTACTGGTCCCAATTTTGGCTATATCAAACTTAGTCCA 304
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: April 27, 2006, 00:22:15  
Job time : 612.88 secs



**THIS PAGE BLANK (U8PT0)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:55:21 ; Search time 25.164 Seconds  
(without alignments)  
1288.550 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVACVQAAPIMYLE.....DLRPYALGKASSETGAQLEEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	836.5	47.4	366	2 A45070	nitrilase - Rhodoc
2	819	46.4	356	2 A47181	nitrilase (EC 3.5.
3	800	45.3	383	2 A43470	aliphatic nitrilas
4	638	36.1	354	2 JC4212	nitrilase (EC 3.5.
5	634.5	35.9	349	1 A28658	nitrilase (EC 3.5.
6	559	31.7	368	2 JQ1613	cyanide hydratase
7	523.5	29.7	346	2 S77025	nitrilase (EC 3.5.
8	499.5	28.3	346	2 T49147	nitrilase (EC 3.5.
9	493.5	28.0	346	2 S22398	nitrilase (EC 3.5.
10	492.5	27.9	339	2 S31969	nitrilase (EC 3.5.
11	491.5	27.8	339	2 T52259	nitrilase (EC 3.5.
12	490.5	27.8	346	2 T49148	nitrilase (EC 3.5.
13	489.5	27.7	339	2 T52262	nitrilase (EC 3.5.
14	453.5	25.7	305	2 T27679	probable nitrilase
15	451.5	25.6	362	2 T52266	nitrilase-like pro
16	445	25.2	348	2 T03739	nitrilase (EC 3.5.
17	443.5	25.1	355	2 T52265	nitrilase (EC 3.5.
18	427.5	24.2	349	2 T03736	nitrilase (EC 3.5.
19	257.5	14.6	199	2 S50363	nitrilase homolog
20	207	11.7	272	2 T41662	probable nitrilase
21	185	10.5	297	2 F75263	probable nitrilase
22	183	10.4	262	2 C71109	hypothetical prote
23	177	10.0	292	2 C87275	hydroxylase, carbon-
24	172.5	9.8	318	2 T48563	hypothetical prote
25	171	9.7	262	2 C75051	hydroxylase related
26	160	9.1	257	2 C69264	conserved hypothet
27	156.5	8.9	579	2 AH3225	amidohydroxylase [lm
28	152.5	8.6	285	2 S58240	hypothetical prote
29	152.5	8.6	295	2 H82556	beta-alanine synth

30	151.5	8.6	322	2 T38399	probable amidohydr
31	150.5	8.5	576	2 G72277	NH(3)-dependent NA
32	146.5	8.3	267	2 B72408	conserved hypothet
33	143	8.1	291	2 S51459	hypothetical prote
34	143	8.1	294	2 AB0115	probable carbon-ni
35	142.5	8.1	292	2 E64558	conserved hypothet
36	142	8.0	312	2 JW0083	N-carbamyl-D-amino
37	140.5	8.0	294	2 G71949	hypothetical prote
38	138.5	7.8	272	2 B69109	N-carbamoyl-D-amin
39	138	7.8	280	2 T34905	probable hydroxylase
40	136.5	7.7	393	2 S27881	beta-alanine synth
41	136	7.7	298	2 T17568	hydroxylase homolog
42	135.5	7.7	290	2 B81369	probable hydroxylase
43	134	7.6	220	2 A84673	probable nitrilase
44	134	7.6	276	2 T40601	putative nitrilase
45	132.5	7.5	292	2 G83608	probable hydratase

ALIGNMENTS

RESULT 1					
A45070					
nitrilase - Rhodococcus rhodochrous					
C/Species: Rhodococcus rhodochrous					
C/Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2004					
C/Accession: A45070					
R/Kobayashi, M.; Komeda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.					
J. Biol. Chem. 267, 20746-20751, 1992					
A/Title: Nitrilase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the					
A/Reference number: A45070; MUID:93015976; PMID:1400390					
A/Contents: J1					
A/Accession: A45070					
A/Status: preliminary					
A/Molecule type: DNA; protein					
A/Residues: 1-366 <KOB>					
A/Cross-references: UNIPROT:Q03217; UNIPARC:UPI000016FF36; GB:D11425; NID:g216933; PIDN:i					
A/Note: sequence extracted from NCBI backbone (NCBIN:116039, NCBIPI:116040)					
C/Superfamily: nitrilase (carbon-nitrogen hydroxylase)					
Query Match					
Best Local Similarity 47.4%; Score 836.5; DB 2; Length 366;					
Matches 156; Conservative 56; Mismatches 95; Indels 1; Gaps 1;					
QY	6	KVACVQAAPIMYLEATVDKTIEMEEARNNARLIAFPETWIPGYWFLWLDSPAWAM- 64			
DB	9	KVAAPVQAPVWFDAKTVDKTIVSIIAARNGCELVAFPEVFIPIGYHIWDSPLAGMA 68			
QY	65	QFVROYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGDTIGA 124			
DB	69	KFAVRHYENSLTMDSPHVQRLLDARDHNIAVVVGISERDGSGLYMTQLVIDADQGLVAR 128			
QY	125	RRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEIHCAA 184			
DB	129	RRKLKPTHYERSVYGEENGSDISYDMPFARLGALNCWEHFQTLTKYAMYSMHEQVHVAS 188			
QY	185	WPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFLVASCALVQSOMIDMLCTDEKHALL 244			
DB	189	WPGMSLYQPEVPAPFVDAQLTATRYALBEGQTFVCTTQVVTPEAHEFFCDNDEQRKLIG 248			
QY	245	AGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDITRLI 304			
DB	249	RGGGFARIIGPDGRDLATPLAEDEGILYADIDLSAITLAKQAADPVGHYSRPDLSLNF 308			
QY	305	DRSPKLPV 312			
DB	309	NQRHTTPV 316			
RESULT 2					
A47181					
nitrilase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis					
C/Species: Alcaligenes faecalis					









A;Cross-references: UNIPROT:P32962; UNIPARC:UPI0000001444; EMBL:X68305; NID:g22655; PID: R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25017  
A;Accession: T49146  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-339 <DAN>  
A;Cross-references: UNIPARC:UPI0000001444; EMBL:AL353865; GSPDB:GN00061; ATSP:T10D17.90  
A;Experimental source: cultivar Columbia; BAC clone T10D17  
R;artling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6021-5, 1994  
A;Title: Molecular characterization of two cloned nitrilases from Arabidopsis thaliana:  
A;Reference number: Z24514; MUID:8016109; PMID:8016109  
A;Accession: T52258  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-339 <BA2>  
A;Cross-references: UNIPARC:UPI0000001444; EMBL:X68305; PIDN:CAA48377.1  
A;Experimental source: cultivar Landsberg erecta  
C;Genetics:  
A;Gene: ATSP:T10D17.90  
A;Map position: 3  
A;Introns: 37/1; 97/1; 195/1; 289/1  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase

Query Match 27.9%; Score 492.5; DB 2; Length 339;  
Best Local Similarity 33.1%; Pred. No. 5.8e-37;  
Matches 112; Conservative 61; Mismatches 110; Indels 55; Gaps 7;

QY 5 IKYACVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPMFLWLDSPAWAM 64  
Db 18 VRATIVQASTVYNDTPATLEKANKFIVEAASKGSELVFPFAFIGYP-----RGF 68  
QY 65 QF-----VRQYHNSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYI 110  
Db 69 RFGLGCVHNEEGRDEFRKYHASAIKVPGEVEKLAELAGKNNVYLVMGAIEKDGTYLYC 128  
QY 111 SQWFIGDNGDTIGARRKLPFTVERTLFGEGDSSLAVFETSVGRGLGCCWEHLQPLTK 170  
Db 129 TALFFSPQGQFLGKHKRLMPTSLERCICWGQDGTIPYDTPIGKLGAAICWENRMDLYR 188  
QY 171 VALYAQNEEIHCA-----AWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFYLASCA 223  
Db 189 TALYAKGIELYCAPTADGSKEQSSMLH-----IAIEGCGFVLSACQ 230  
QY 224 LVSQSMI----DMLCTD--DEKHALLLAGGHSRIIGPDGDLVAPLAENEGILYANLD 277  
Db 231 FCLRKDFPDHPDYLFTDWYDDKEPDSIVSQGSVITISPLGQVLAGPNFES-EGLTADLD 289  
QY 278 PGVRILAKMAADPAGHYSRPDITRLLIDRSKPLPYVEI 315  
Db 290 LGDVARAKLYFDXVGHYSRPDVHLTVNEHPKKPVTFI 327

RESULT 11  
T52259  
nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C;Accession: T52259  
R;Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A;Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A;Reference number: Z26007; MUID:98145459; PMID:9484465  
A;Accession: T52259  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-339 <HIL>  
A;Cross-references: UNIPROT:O04907; UNIPARC:UPI00000A5899; EMBL:Y07648; PIDN:CAA68934.3  
A;Experimental source: cultivar Columbia  
C;Genetics:

A;Gene: nit2  
A;Introns: 37/1; 97/1; 195/1; 289/1  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C;Keywords: hydrolase

Query Match 27.8%; Score 491.5; DB 2; Length 339;  
Best Local Similarity 33.1%; Pred. No. 7.1e-37;  
Matches 112; Conservative 61; Mismatches 110; Indels 55; Gaps 7;

QY 5 IKYACVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPMFLWLDSPAWAM 64  
Db 18 VRATIVQASTVYNDTPATLEKANKFIVEAASKGSELVFPFAFIGYP-----RGF 68  
QY 65 QF-----VRQYHNSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYI 110  
Db 69 RFGLGCVHNEEGRDEFRKYHASAIKVPGEVEKLAELAGKNNVYLVMGAIEKDGTYLYC 128  
QY 111 SQWFIGDNGDTIGARRKLPFTVERTLFGEGDSSLAVFETSVGRGLGCCWEHLQPLTK 170  
Db 129 TALFFSPQGQFLGKHKRLMPTSLERCICWGQDGTIPYDTPIGKLGAAICWENRMDLYR 188  
QY 171 VALYAQNEEIHCA-----AWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFYLASCA 223  
Db 189 TALYAKGIELYCAPTADGSKEQSSMLH-----IAIEGCGFVLSACQ 230  
QY 224 LVSQSMI----DMLCTD--DEKHALLLAGGHSRIIGPDGDLVAPLAENEGILYANLD 277  
Db 231 FCLRKDFPDHPDYLFTDWYDDKEPDSIVSQGSVITISPLGQVLAGPNFES-EGLTADLD 289  
QY 278 PGVRILAKMAADPAGHYSRPDITRLLIDRSKPLPYVEI 315  
Db 290 LGDVARAKLYFDXVGHYSRPDVHLTVNEHPKKPVTFI 327

RESULT 12  
T49148  
nitrilase (EC 3.5.5.1) 3 [imported] - Arabidopsis thaliana  
N;Alternate names: protein T10D17.110  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 31-Dec-2004  
C;Accession: T49148; T52261; T52264  
R;D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25017  
A;Accession: T49148  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-346 <DAN>  
A;Cross-references: UNIPROT:P46010; UNIPARC:UPI000000BDB6; EMBL:AL353865; GSPDB:GN00061;  
A;Experimental source: cultivar Columbia; BAC clone T10D17  
R;Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A;Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A;Reference number: Z26007; MUID:98145459; PMID:9484465  
A;Accession: T52261  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-346 <HIL>  
A;Cross-references: UNIPARC:UPI000000BDB6; EMBL:Y07648; PIDN:CAA68936.2  
A;Experimental source: cultivar Columbia  
R;Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A;Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A;Reference number: Z24515; MUID:8022831; PMID:8022831  
A;Accession: T52264  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-346 <BAR>  
A;Cross-references: UNIPARC:UPI000000BDB6; EMBL:U09959; PIDN:AAA19627.1  
A;Experimental source: cultivar Columbia  
C;Genetics:  
A;Gene: ATSP:T10D17.110; nit3  
A;Map position: 3



A;Accession: T52266  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-362 <CHI>  
A;Cross-references: UNIPROT:Q9SXX6; UNIPARC:UPI000009C92A; EMBL:AB027054; PIDN:BAA77679.  
A;Experimental source: cultivar Nipponbare  
C;Genetics:  
A;Gene: ONIT4  
C;Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 25.6%; Score 451.5; DB 2; Length 362;  
Best Local Similarity 33.5%; Pred. No. 3.4e-33;  
Matches 109; Conservative 62; Mismatches 123; Indels 31; Gaps 7;

QY 5 IKVACVQAPIYMDLEATVDKTIEMEEAARNNARLIAPPETWIPGYP-----WFLW 56  
Db 31 VRATVVQASTVFYDTPATLDKAERLIEAAGYGSQLVFPEAFVGYPFGSTFGANIS 90  
QY 57 LDSPA-WAMQFVROYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGTLYISQWFI 115  
Db 91 IGNPKDKGKEBFKRYHAAIEVPGEVTRILAAAMAGKYKFLVMGVIEREGYTLVCSVLF 150  
QY 116 GDNGDTIGARRKLTPTFVERTLFGEGDSSLAVFETSVGRLGGLCCWEHLQPLTKYALYA 175  
Db 151 DPLGRYLGKHKRLMPTALERLIWFGDGSTIPYDTPLGKIGALLICWENKMPLRLTALYG 210  
QY 176 QNEEIHCAAMPSSFSLYPNAKALGPDVNVASRIYAVEGQCFVLASCALVSQS----- 228  
Db 211 KGIEIYCA-----PTADSR---QVWQASMTHTALEGGCFVLSANQFCRRKDYPPPE 259  
QY 229 -MIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLYANLDPGVRIIAKMA 287  
Db 260 YVFTGLGEPSPDIVVCPGG--SVIISPSGEVLGPNYEG-EALITADLDLGEIVRAKFD 316  
QY 288 ADPAGHYSRPDITRLLIDRSPKLPV 312  
Db 317 FDVVGHYARPEVLSLVVNDQPHLPV 341

Search completed: April 27, 2006, 01:01:07  
Job time : 25.664 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:35 ; Search time 151.477 Seconds  
(without alignments)  
1569.628 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVACVQAPITMDLE.....DLRPYALGKASETGAQLEEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1765	100.0	337	2 Q6RWE0_9ZZZZZ	Q6rwe0 uncultured
2	1343	76.1	337	2 Q6RWQ3_9ZZZZZ	Q6rww3 uncultured
3	1343	76.1	337	2 Q6RWQ7_9ZZZZZ	Q6rww7 uncultured
4	1187	67.3	336	2 Q500U1_PSESY	Q500u1 pseudomonas
5	1184	67.1	347	2 Q88B32_PSESM	Q88b32 pseudomonas
6	1139	64.5	333	2 Q6RWL0_9ZZZZZ	Q6rwl0 uncultured
7	1136	64.4	332	2 Q6RWU0_9ZZZZZ	Q6rwj0 uncultured
8	1081	61.2	345	2 Q6RWS5_9ZZZZZ	Q6rws5 uncultured
9	1051	59.5	334	2 Q6RWE3_9ZZZZZ	Q6rwe3 uncultured
10	1042	59.0	353	2 Q6RWP9_9ZZZZZ	Q6rwp9 uncultured
11	1039	58.9	348	2 Q6RWS2_9ZZZZZ	Q6rws2 uncultured
12	1037	58.8	336	2 Q6RWE6_9ZZZZZ	Q6rwe6 uncultured
13	1037	58.8	353	2 Q6RWF0_9ZZZZZ	Q6rwf0 uncultured
14	1033	58.5	337	2 Q6RWS3_9ZZZZZ	Q6rws3 uncultured
15	1033	58.5	338	2 Q6RWE4_9ZZZZZ	Q6rwe4 uncultured
16	1028.5	58.3	358	2 Q6RWI0_9ZZZZZ	Q6rwi0 uncultured
17	1022.5	57.9	337	2 Q6RWN8_9ZZZZZ	Q6rwn8 uncultured
18	1021.5	57.9	337	2 Q6RWN9_9ZZZZZ	Q6rwn9 uncultured
19	1020.5	57.8	354	2 Q6RWE6_9ZZZZZ	Q6rwe6 uncultured
20	1019.5	57.8	338	2 Q6RWG6_9ZZZZZ	Q6rww6 uncultured
21	1018	57.7	338	2 Q6RWI7_9ZZZZZ	Q6rwi7 uncultured
22	1008	57.1	337	2 Q6RWS3_9ZZZZZ	Q6rws3 uncultured
23	1007	57.1	346	2 Q6RWS5_9ZZZZZ	Q6rws5 uncultured
24	999.5	56.6	338	2 Q6RWN2_9ZZZZZ	Q6rwn2 uncultured
25	999	56.6	338	2 Q6RWN6_9ZZZZZ	Q6rwn6 uncultured
26	997.5	56.5	336	2 Q5YUM5_NOCFA	Q5yum5 nocardia fa
27	994.5	56.3	338	2 Q6RWQ4_9ZZZZZ	Q6rww4 uncultured
28	993	56.3	338	2 Q6RWH1_9ZZZZZ	Q6rwh1 uncultured
29	980.5	55.6	327	2 Q6RWH9_9ZZZZZ	Q6rwh9 uncultured
30	976	55.3	352	2 Q6RWH0_9ZZZZZ	Q6rwh0 uncultured
31	973.5	55.2	336	2 Q6RWJ5_9ZZZZZ	Q6rwj5 uncultured

32	971.5	55.0	341	2 Q6RWR4_9ZZZZZ	Q6rwr4 uncultured
33	937.5	53.1	335	2 Q6RWF7_9ZZZZZ	Q6rwf7 uncultured
34	936	53.0	331	2 Q706Q8_PSEPU	Q706q8 pseudomonas
35	929.5	52.7	337	2 Q6RWM3_9ZZZZZ	Q6rwm3 uncultured
36	924	52.4	350	2 Q5EG61_PSEFL	Q5eg61 pseudomonas
37	919.5	52.1	334	2 Q6RWG9_9ZZZZZ	Q6rww9 uncultured
38	918.5	52.0	334	2 Q89GB3_BRAJA	Q89ge3 bradyrhizob
39	914.5	51.8	337	2 Q6RWM9_9ZZZZZ	Q6rwm9 uncultured
40	906	51.3	345	2 Q6RWN2_9ZZZZZ	Q6rwn2 uncultured
41	898.5	50.9	349	2 Q6RWS0_9ZZZZZ	Q6rws0 uncultured
42	881	49.9	346	2 Q6RWM1_9ZZZZZ	Q6rwm1 uncultured
43	836.5	47.4	365	1 NRL2_RHORH	Q03217 rhodococcus
44	819	46.4	356	1 NRLA_ALCFA	P20960 alcaligenes
45	800	45.3	383	1 NRL1_RHORH	Q02068 rhodococcus

ALIGNMENTS

RESULT 1									
ID	Q6RWE0_9ZZZZZ	PRELIMINARY;	PRT;	337	AA.				
AC	Q6RWE0;								
DT	05-JUL-2004 (Tremblrel. 27, Created)								
DT	05-JUL-2004 (Tremblrel. 27, Last sequence update)								
DT	05-JUL-2004 (Tremblrel. 27, Last annotation update)								
DE	Nitrilase (EC 3.5.5.7).								
GN	ORFNames=BD5220;								
OS	uncultured organism.								
OC	unclassified; environmental samples.								
OX	NCBI_TaxId=155900;								
RN	[1]								
RP	NUCLEOTIDE SEQUENCE.								
RX	PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;								
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,								
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,								
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,								
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.;								
RT	"Exploring nitrilase sequence space for enantioselective catalysis.";								
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).								
DR	EMBL; AY487562; AAR97509.1; -; Genomic DNA.								
DR	GO; GO:0018762; F:aliphatic nitrilase activity; IEA.								
DR	GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.								
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.								
DR	InterPro; IPR0010132; N/CN hydrtase.								
DR	InterPro; IPR003010; Ntlse/CNhydrtse.								
DR	Pfam; PF00795; CN_hydrolyase; 1.								
DR	PROSITE; PS50263; CN HYDROLASE; 1.								
DR	PROSITE; PS00921; NITRIL_CHT_2; 1.								
KW	Hydrolyase.								
SQ	SEQUENCE 337 AA; 36783 MW; 25D2DE1E820983B2 CRC64;								
Query Match 100.0%; Score 1765; DB 2; Length 337;									
Best Local Similarity 100.0%; Pred. No. 4.9e-145;									
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps									
QY	1 MKEAIKVACVQAPITMDLEATVDKTIELMEBAARNNA RLIAFPETWIPGYPWFLMDSP 60								
DB	1 MKEAIKVACVQAPITMDLEATVDKTIELMEBAARNNA RLIAFPETWIPGYPWFLMDSP 60								
QY	61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQDNGD 12								
DB	61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQDNGD 12								
QY	121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGGLCCWEHLQPLTKYALYAQNEEI 18								
DB	121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGGLCCWEHLQPLTKYALYAQNEEI 18								
QY	181 HCAAMPFSFLYPNAKALGPDVNVAA SRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 24								
DB	181 HCAAMPFSFLYPNAKALGPDVNVAA SRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 24								
QY	241 ALLLAGGHSRIIGPDGDLVAPLAENEGLYANLDPGVRI LAKMAADPAGHYSRPDIT 300								

DB 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRI LAKMAADPAGHSRPDIT 300  
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337

RESULT 2  
Q6RMQ3 9ZZZZ PRELIMINARY; PRT; 337 AA.

AC Q6RMQ3: 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5275;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RA "Exploring nitrilase sequence space for enantioselective catalysis.";  
RT Appl. Environ. Microbiol. 70:2429-2436(2004).  
RL EMBL; AY487449; AAR97396.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR00132; N/CN\_hydrolase.  
DR InterPro; IPR003010; Ntase/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolyase.  
SQ SEQUENCE 337 AA; 36697 MW; 940993D5AB69582E CRC64;

Query Match 76.1%; Score 1343; DB 2; Length 337;  
Best Local Similarity 75.5%; Pred. No. 2.7e-108;  
Matches 253; Conservative 29; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVACVQAAPYIMDL EATVDKTIEMEEAARNNARLIAFPETWIPGYPFWLWLDSP 60  
DB 1 MKEAIKVACVQAAPVFLDL DATVDKTVALLIEAARNGARLIAFPETWIPGYPFWLWLDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLIYSQWFIGDNGD 120  
DB 61 AWGMOFVRRYHENS LVDSPQAKRIS EAAQRA GITYVALGYSERVS GTLYMGQWLIDDKGE 120  
QY 121 TIGARRKCLKPTFVERTLFGEGDSSSLAVFETSVGR LGGLCCWEHLQPLTKYALYAQNEEI 180  
DB 121 TAGLRKCLKPTHVERTLFGEGDSSSLSTFTPLGVLGGLCCWEHLQPLSKYALYAQNEEI 180  
QY 181 HCAAMPFSFLYPNAKALGPDVNVAA SRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HFAAMPFSFIYRQATEVILGPEVNVAASRIYAVEGQCFVLASCALVSPENIEMLTCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRI LAKMAADPAGHSRPDIT 300  
DB 241 SLQAGGYSRIIGPDGSDLARPLGENEGILYATLDPAR IYAKTAADPAGHSRPDVT 300  
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLE 335  
DB 301 RLINRSANQPVVEVGREI-----PASAQGFVEV 329

RESULT 3  
Q6RMQ7\_9ZZZZ

ID Q6RMQ7\_9ZZZZ PRELIMINARY; PRT; 337 AA.  
AC Q6RMQ7: 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5274;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RA "Exploring nitrilase sequence space for enantioselective catalysis.";  
RT Appl. Environ. Microbiol. 70:2429-2436(2004).  
RL EMBL; AY487445; AAR97392.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR00132; N/CN\_hydrolase.  
DR InterPro; IPR003010; Ntase/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolyase.  
SQ SEQUENCE 337 AA; 36769 MW; 940990B5AB69582E CRC64;

Query Match 76.1%; Score 1343; DB 2; Length 337;  
Best Local Similarity 75.5%; Pred. No. 2.7e-108;  
Matches 253; Conservative 29; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVACVQAAPYIMDL EATVDKTIEMEEAARNNARLIAFPETWIPGYPFWLWLDSP 60  
DB 1 MKEAIKVACVQAAPVFLDL DATVDKTVALLIEAARNGARLIAFPETWIPGYPFWLWLDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLIYSQWFIGDNGD 120  
DB 61 AWGMOFVRRYHENS LVDSPQAKRIS EAAQRA GITYVALGYSERVS GTLYMGQWLIDDKGE 120  
QY 121 TIGARRKCLKPTFVERTLFGEGDSSSLAVFETSVGR LGGLCCWEHLQPLTKYALYAQNEEI 180  
DB 121 TAGLRKCLKPTHVERTLFGEGDSSSLSTFTPLGVLGGLCCWEHLQPLSKYALYAQNEEI 180  
QY 181 HCAAMPFSFLYPNAKALGPDVNVAA SRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HFAAMPFSFIYRQATEVILGPEVNVAASRIYAVEGQCFVLASCALVSPENIEMLTCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRI LAKMAADPAGHSRPDIT 300  
DB 241 SLQAGGYSRIIGPDGSDLARPLGENEGILYATLDPAR IYAKTAADPAGHSRPDVT 300  
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLE 335  
DB 301 RLINRSANQPVVEVGREI-----PASAQGFVEV 329

RESULT 4  
Q500U1\_PSESY PRELIMINARY; PRT; 336 AA.  
ID Q500U1\_PSESY  
AC Q500U1: 13-SEP-2005 (Tremblrel. 31, Created)  
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)  
DT 13-SEP-2005 (Tremblrel. 31, Last annotation update)  
DE Aliphatic nitrilase (EC 3.5.5.7).  
GN ORFNames=Psyr\_0007;  
OS Pseudomonas syringae pv. syringae B728a.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=205918;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RG DOE Joint Genome Institute;  
RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,  
RA Nolan M., Goldsman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,  
RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.;  
RT "Comparison of two complete genome sequences of *Pseudomonas syringae*  
RT pv. *syringae* B728a and pv. *tomato* DC3000.";  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RA Loper J.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RA Feil H., Feil W.S., Lindow S.E.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; CP000075; AAY35081.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016787; F:hydrolase activity; IEA.  
KW Hydrolase.  
SQ SEQUENCE 336 AA; 36552 MW; EC3DF2A762E37729 CRC64;

Query Match 67.3%; Score 1187; DB 2; Length 336;  
Best Local Similarity 69.2%; Pred. No. 1.1e-94;  
Matches 218; Conservative 37; Mismatches 60; Indels 0; Gaps 0;

QY 1 MKEAIKVCQVQAPPIYMDLEATVDKTIELMEEAARNRRLIAPPETWIPGYPWLMDSP 60  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 1 MKEPLKVCQVQAPVFLDLATVDKTIELMEQAAAAGAGLIAFPETWIPGYPWLMDAP 60  
QY 61 AWAMQFVROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLYISQWFIGDNGD 120  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 61 AMNMPLVQRHYQOSLVLDVQARRISDAARHLGLVVLGYSERNKASLYIGQWIIDHGE 120  
QY 121 TIGARRKLTPTFVERTLFGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 121 TVGVRKRLKATHVERTLFGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYAMYAQNEOI 180  
QY 181 HCAAMPSSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 181 HVAAMPSSFLYRNATSGLEVENTAASRYVAEGQCFVLAPCAIVSPENIEMLCDSDAKR 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300  
:|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 241 SLQAGGSHARIFGPDGSDLATPLGEHEEGLLYATLDPALTLAKVAADPAGHSRPDVT 300  
QY 301 RLIDRSPKLPVVEI 315  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 301 RLMEFNPNPTPCVVDL 315

RESULT 5  
Q88B32\_PSESM  
ID Q88B32\_PSESM PRELIMINARY; PRT; 347 AA.  
AC Q88B32;  
DT 01-JUN-2003 (Tremblrel. 24, Created)  
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Nitrilase, putative.  
GN OrderedLocusNames=PSPT00189;  
OS *Pseudomonas syringae* (pv. *tomato*).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=DC3000;  
MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;

RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,  
RA Gwin M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,  
RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,  
RA Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,  
RA Uterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinour S., Chatterjee A.K.,  
RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,  
RA Bender C.L., White O., Fraser C.M., Collier A.;  
RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
RT *Pseudomonas syringae* pv. *tomato* DC3000.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
DR EMBL; AB016853; AAO53742.1; -; Genomic\_DNA.  
DR TIGR; PSPT00189; -;  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrolase.  
DR InterPro; IPR003010; Ntlase/CNhydrtase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 347 AA; 37564 MW; B671F783D7411308 CRC64;

Query Match 67.1%; Score 1184; DB 2; Length 347;  
Best Local Similarity 68.8%; Pred. No. 2e-94;  
Matches 218; Conservative 37; Mismatches 62; Indels 0; Gaps 0;

QY 1 MKEAIKVCQVQAPPIYMDLEATVDKTIELMEEAARNRRLIAPPETWIPGYPWLMDSP 60  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 12 MKEPLKVCQVQAPVFLDLATVDKTIELMEQAAAAGAGLIAFPETWIPGYPWLMDAP 71  
QY 61 AWAMQFVROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLYISQWFIGDNGD 120  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 72 AMNMPLVQRHYQOSLVLDVQARRISDAARHGLVVLGYSERNKASLYIGQWIIDHGE 131  
QY 121 TIGARRKLTPTFVERTLFGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 132 TVGTRKRLKATHVERTLFGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYAMYAQNEOI 191  
QY 181 HCAAMPSSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 192 HVAAMPSSFLYQNATRALGPEVENTAASRYVAEGQCFVLAPCAIVSPENIEMLCDSDAKH 251  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300  
:|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 252 SLQAGGSHARIFGPDGSDLATPLGEHEEGLLYATLDPALTLAKVAADPAGHSRPDVT 311  
QY 301 RLIDRSPKLPVVEIEG 317  
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|  
Db 312 RLMEFNPNPCVVELPG 328

RESULT 6  
Q6RWLO\_9ZZZZ  
ID Q6RWLO\_9ZZZZ PRELIMINARY; PRT; 333 AA.  
AC Q6RWLO;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7615;  
OS unclassified, environmental samples.  
OC unclassified, environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;



```
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487492; AACR97439.1; -; Genomic_DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydrtase.
DR InterPro; IPR003010; Ntlse/CNhydttse.
DR Pfam; PF00795; CN_hydrolyase; 1.
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW Hydrolyase.
SQ SEQUENCE 333 AA; 36342 MW; B0A9B825FD0D7403 CRC64;

Query Match 64.5%; Score 1139; DB 2; Length 333;
Best Local Similarity 67.5%; Pred. No. 1.6e-90;
Matches 212; Conservative 36; Mismatches 66; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAAPRYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPFWFLWDSP 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2 MKTTVTVACVQAAPVFMDELTIDKTIITLISEAAQGAELIAFPETWIPGYPFWFLWNSP 61

QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTILGMSERVGTLYISQWFIGDNGD 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 ATNMPLVYQYHONS�VLDSTQAKRIADAARQNNITVALGFSERDHGS�YIAQWLIGSDGE 121

QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 122 TIGIRRLKATHVERTLFGESDSSLTWTETPLGNVGALCCWEHLQPLSRYAMYSQHEEI 181

QY 181 HCAAMPSESLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 182 HIAAMPSESLYTSATALGPDVNTAASRLYAAEGQCFVLAPCAVVSDEMIDLCPPDDRR 241

QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEBGILYANLDPGVRILAKMAADPAGHSRPDIT 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 242 ALLSAGGHARIYGPDGRELVTPLGENEBGLLIAELDSAATITFAKLAADPVGHYSRPDVT 301

QY 301 RLIDRSPKLPVE 314
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 302 RLLENPSANKTVIK 315
```

```
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW Hydrolyase.
SQ SEQUENCE 332 AA; 36214 MW; 2E77B8470F426184 CRC64;

Query Match 64.4%; Score 1136; DB 2; Length 332;
Best Local Similarity 67.2%; Pred. No. 2.9e-90;
Matches 211; Conservative 37; Mismatches 66; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAAPRYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPFWFLWDSP 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKTTVTVACVQAAPVFMDELTIDKTIITLISEAAQGAELIAFPETWIPGYPFWFLWNSP 60

QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTILGMSERVGTLYISQWFIGDNGD 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 ATNMPLVYQYHONS�VLDSTQAKRIADAARQNNITVALGFSERDHGS�YIAQWLIGSDGE 120

QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 TIGIRRLKATHVERTLFGESDSSLTWTETPLGNVGALCCWEHLQPLSRYAMYSQHEEI 180

QY 181 HCAAMPSESLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HIAAMPSESLYTSATALGPDVNTAASRLYAAEGQCFVLAPCAVVSDEMIDLCPPDDRR 240

QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEBGILYANLDPGVRILAKMAADPAGHSRPDIT 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 ALLSAGGHARIYGPDGRELVTPLGENEBGLLIAELDSAATITFAKLAADPVGHYSRPDVT 300

QY 301 RLIDRSPKLPVE 314
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 RLLENPSANKTVIK 314

RESULT 8
Q6RWS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
ID Q6RWS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
AC Q6RWS5;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD7611;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487427; AAR97374.1; -; Genomic_DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydratase.
DR InterPro; IPR003010; Ntlse/CNhydttse.
DR Pfam; PF00795; CN_hydrolyase; 1.
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW Hydrolyase.
SQ SEQUENCE 345 AA; 37242 MW; 5F8A3CFF530D2FID CRC64;

Query Match 61.2%; Score 1081; DB 2; Length 345;
Best Local Similarity 65.3%; Pred. No. 1.8e-85;
Matches 201; Conservative 41; Mismatches 66; Indels 0; Gaps 0;

QY 6 KVAACVQAAPRYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPFWFLWDSPAMQ 65
```

Db 9 RVAAVQAPEFLNLEATVDTKTIALIEBARAGSALIAFPETWIPGYPMFAMLGAPIWGMK 68  
QY 66 FVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGDTIGAR 125  
Db 69 FIOAYHDNSMVIDGAQOFERIAQAASRCNITTVLGFSEKDGASLYIAQAILSPGKTIATR 128  
QY 126 RKLKPTFVERTLFEGDGSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEIHCAM 185  
Db 129 RKLKPTHVERAIFEGGDSDLAVHDTKGRVGALCCWEHLQPLSKYAMYAQNEQVHIAAW 188  
QY 186 PSFSLYPNAKALGPDVNVASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALLLA 245  
Db 189 PSFSLYVDAAAYALGPEVNNAASRLYAVEGQCFVAPCATVSQKMIDMLCETPEQOALLKP 248  
QY 246 GGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHYSRPDTIRLLID 305  
Db 249 GGGHAQIYGPDGSLADPLPPDAEGLLYADIDLAAITLAKAAADPAGHYSRPDTIQLLLD 308  
QY 306 RSPKLPV 313  
Db 309 RNPKPRV 316

RESULT 9

Q6RWE3\_9Z2ZZZ PRELIMINARY; PRT; 334 AA.

AC Q6RWE3;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrlase (EC 3.5.5.7).  
GN ORFNames=BD7051;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrlase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487559; AAR97506.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrlase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydts.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 334 AA; 36445 MW; AOB271DDFF33735E CRC64;

Query Match 59.5%; Score 1051; DB 2; Length 334;  
Best Local Similarity 60.7%; Pred. No. 7.3e-83;  
Matches 199; Conservative 51; Mismatches 68; Indels 10; Gaps 3;

QY 4 AIKVACVQAPIYMDLEATVDTKTIELMEBARNNARLIAFPETWIPGYPMFLWLDSPAWA 63  
Db 7 AFKAADVQAPEWLDLDRITVDKTIALIEBAAGAKLIAFPETWIPGYPMHIVGTPTAWA 66  
QY 64 MQ--FVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGDT 121  
Db 67 ISRGFVQRYFDNSLAYSPQAQRIADAACKNKITVVLGLSEREGSLTYSQWMLIGPDGET 126  
QY 122 IGARRKLKPTFVERTLFEGDGSSLA VFE-TSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180  
Db 127 IAKRRKLKPTHVERTVFGDGSGSHIAVHERADIGRLGALCCWEHLQPLTKYAMYAQNEQV 186

QY 181 HCAAMPFSLYPNAKALGPDVNVASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 187 HVAAMPFSPSMYEPFAHALGWEVNNAASKIYAVEGSCFVLGACAVISQAMVDEMCDTEDKR 246  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHYSRPDT 300  
Db 247 ALVHAGGHAIVFGPDGRSLADKIPETQEGLLYADIDLGAIGVAKNAADPAGHYSRPDT 306  
QY 301 RLIDRSP-----KLPEVIEGDLRP 321  
Db 307 RLLENKKPARREVEYFSLPVDVAVETPPQP 334

RESULT 10

Q6RMF9\_9Z2ZZZ PRELIMINARY; PRT; 353 AA.

AC Q6RMF9;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitrlase (EC 3.5.5.7).  
GN ORFNames=BD7753;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrlase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487543; AAR97490.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrlase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydts.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC31326 CRC64;

Query Match 59.0%; Score 1042; DB 2; Length 353;  
Best Local Similarity 63.6%; Pred. No. 4.7e-82;  
Matches 194; Conservative 44; Mismatches 65; Indels 2; Gaps 1;

QY 6 KVAACVQAAPITMDLEATVDTKTIELMEBARNNARLIAFPETWIPGYPMFLWLDSPAWAQ 65  
Db 9 KVAADVQAPEFLDLDAVEKAVRFIDEAGAAARLIAFPETWIPGYPMIWLGAFAWAIM 68  
QY 66 --FVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGDTIG 123  
Db 69 RGFVSRYPFNSLSYDSPQAEKLRRAAKRMVVVLGLSERDGSLTIAQMIIGPDGETIA 128  
QY 124 ARRKLPPTFVERTLFEGDGSSLA VETS VGR LGGLCCWEHLQPLTKYALYAQNEIHC 183  
Db 129 KRRKLKPTAERTVFEEDGSHLAVHELDVRLGALCCWEHLQPLSKYAMYAQNEQVHVA 188  
QY 184 AMPSFSLYPNAKALGPDVNVASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALL 243  
Db 189 AMPSFSLYDPEFAHALGABVNNAASKIYAVEGSCFVIAPCATVSQAMIDELCTPEKHQFL 248  
QY 244 LAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHYSRPDTIRLL 303  
Db 249 HAAGGFAVYIGPDGAPLAAPLPDKGGLLYADIDLGMISVAKAADPAGHYARPDTIRLL 308  
QY 304 IDRSP 308

Db 309 FNNRP 313

RESULT 11  
Q6RWS2\_9ZZZZ PRELIMINARY; PRT; 348 AA.  
ID Q6RWS2\_9ZZZZ  
AC Q6RWS2;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitriase (EC 3.5.5.7).  
GN ORFNames=BD5338;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487430; AAR97377.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitriase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydase.  
DR InterPro; IPR002332; PII\_GlnB\_UMP\_S.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
DR PROSITE; PS00496; PII\_GlnB\_UMP; UNKNOWN\_1.  
KW Hydrolase.  
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;

Query Match 58.9%; Score 1039; DB 2; length 348;  
Best Local Similarity 63.0%; Pred. No. 8.5e-82;  
Matches 191; Conservative 46; Mismatches 66; Indels 0; Gaps 0;  
QY 6 KVAACVQAAPITYMDLEATVDKTIELMEBAARNARLIAFPETWIPGYPFWLWLDSPAWAQ 65  
Db 8 RVAAVQAAPVFLDLEGAISKGISLIEEASNGAKLIAFPETWIPGYPFWLWLDSPAWGMR 67  
QY 66 FVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTISQWFIGDNGDTIGAR 125  
Db 68 FVQRYFDNSLMLGSEQAKRMNQAAANNKIYVVMGYSERSGSLYMGQSIINDKGETIFTR 127  
QY 126 RKLKPTFEVERTLFGEGDSSLAVFETSVGRLLGCLCWEHLQPLTKYALYAQNEEIHCAAW 185  
Db 128 RKLKPTFEVERTVFGEGDSSLTVMDEIGRVGAMCCWEHLQPLSKYAMYSQDEQIHIAW 187  
QY 186 PSFSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALLIA 245  
Db 188 PSFSLYRGAAAYALGPELNNAASQWYAEGQCFVLAPCATVSKEMIEMLIIDPRKEPLILE 247  
QY 246 GGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIAKMAADPAGHYSRPDITRLLID 305  
Db 248 GGGFTMTYGPDGRPLAKPLPENEEGLYADIDLGMISMAKAAADPAGHYARPDVTRLLFN 307  
QY 306 RSP 308  
Db 308 SAP 310

RESULT 12  
Q6RWF6\_9ZZZZ PRELIMINARY; PRT; 336 AA.  
ID Q6RWF6\_9ZZZZ  
AC Q6RWF6;

DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitriase (EC 3.5.5.7).  
GN ORFNames=BD5295;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487546; AAR97493.1; -; Genomic\_DNA.  
DR GO; GO:0018762; F:aliphatic nitriase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN\_hydrtase.  
DR InterPro; IPR003010; Ntlse/CNhydase.  
DR Pfam; PF00795; CN\_hydrolase; 1.  
DR PROSITE; PS50263; CN\_HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
KW Hydrolase.  
SQ SEQUENCE 336 AA; 36221 MW; CC370C188F94CB72 CRC64;

Query Match 58.8%; Score 1037; DB 2; length 336;  
Best Local Similarity 60.1%; Pred. No. 1.2e-81;  
Matches 197; Conservative 46; Mismatches 73; Indels 12; Gaps 3;  
QY 6 KVAACVQAAPITYMDLEATVDKTIELMEBAARNARLIAFPETWIPGYPFWLWLDSPAWAQ 65  
Db 9 KAAVQAAPVFLDLDGVERAVSLIGQAAEGAQLIAFPETWIPGYPFWHTWLGSPAWAME 68  
QY 66 --FVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTISQWFIGDNGDTIG 123  
Db 69 KGFVQRYFDNALRHGSPQAERISGAAEHKIMVSLGFAERDGGTLVIAQWLIGPDQTTIS 128  
QY 124 ARRLKPTFEVERTLFGEGDSSLAVFETSVGRLLGCLCWEHLQPLTKYALYAQNEEIHCA 183  
Db 129 RRRKLKPTFEVERTVFGEGDSSLVHDTALGRIGSLCWEHLQPLSKYAMYAQNEQIHIG 188  
QY 184 AMPFSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALI 243  
Db 189 AMPFSLYQPPANALISPEVNIAVSRVYAVEGQCFVLAPCATVSDAMTETLCDTPEKGLI 248  
QY 244 LAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIAKMAADPAGHYSRPDITRLL 303  
Db 249 RAGGHAALFPGDSSLTPTVADTYEGLIYALDGLVISIAKSAADPAGHYSRPDVTRLL 308  
QY 304 IDRSPK-----LP---VVEIEGDLRP 321  
Db 309 LNQTPSKRVQNWVLPLETVTEPEGPVQP 336

RESULT 13  
Q6RWF0\_9ZZZZ PRELIMINARY; PRT; 353 AA.  
ID Q6RWF0\_9ZZZZ  
AC Q6RWF0;  
DT 05-JUL-2004 (Tremblrel. 27, Created)  
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
DE Nitriase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;





Best Local Similarity 60.1%; Pred. No. 2.7e-81;  
Matches 194; Conservative 53; Mismatches 66; Indels 10; Gaps 3;

```
QY      6 KVACVQAPITMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPMFLWLDSPAWAM- 64
      :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      9 RVAVVQAPAWLDEASVSKSIALIEEAAAKGAKLIAFPEAFIPGYPMYIWLDSPAWAIG 68

QY      65 -QFVRQYHENSLELDGPOAKRISDAKRLGIMWTLGMSERVGGLYISQWFIQDNGDTIG 123
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      69 RGFVQRYFDNSLSYDSFOAERLRILAVKKAGMTAVLGISERDGSLYLAQWILGPDGETIA 128

QY      124 ARRKLPFTFVERTLEFGEQDSSSLAVFE-TSVGRIGLICWEHLQPLTKYALYAQNEEIH 182
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      129 KRRKLPTHAERTVYGEQDSDLAVHDPGIGRLGALCCWEHLQPLSKYAMYAONEQVHV 188

QY      183 AAWPSFSLYPNAKALGPDVNVVAASRIYAVEGQCFVILASCALVSQSMIDMLCTDDEKHAL 242
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      189 AAWPSFSLYDPFAPALGWEVNNVAASRVYAVEGSCFVILAPCATVSQAMIDELCDRDDKHAL 248

QY      243 LLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHYSRPDITRL 302
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      249 LHVGGGHAALFGPDGSAIADKLPSDQEGLLFADIDLGAIGIAKNADPAGHYSRPDVTRL 308

QY      303 LIDRSPK-----LPVVEIEGD 318
      |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db      309 LNKKPSKRVEHFALPLDTLAGE 331
```

Search completed: April 27, 2006, 01:00:11  
Job time : 152.477 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:05 ; Search time 132.234 Seconds  
(without alignments)  
1119.760 Million cell updates/sec

Title: US-09-751-299-4

Perfect score: 1765

Sequence: 1 MKEAIKVACVQAAPITMDLE.....DLRPYALGKASGTGAQLLEI 337

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_21: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*  
9: geneseqp2005s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765	100.0	337	4	AAE05489	Aae05489 Nitrilase
2	1761	99.8	337	7	ADC24119	Adc24119 Protein s
3	1761	99.8	337	8	ADH36220	Adh36220 Chemical
4	1761	99.8	337	8	ADG93919	Adg93919 Nitrilase
5	1761	99.8	337	8	ADI62518	Adi62518 Nitrilase
6	1761	99.8	337	8	ADI64639	Adi64639 Nitrilase
7	1339	75.9	337	7	ADC23791	Adc23791 Protein s
8	1339	75.9	337	7	ADC23781	Adc23781 Protein s
9	1339	75.9	337	8	ADH35882	Adh35882 Chemical
10	1339	75.9	337	8	ADH35892	Adh35892 Chemical
11	1339	75.9	337	8	ADG93583	Adg93583 Nitrilase
12	1339	75.9	337	8	ADG93593	Adg93593 Nitrilase
13	1339	75.9	337	8	ADI62180	Adi62180 Nitrilase
14	1339	75.9	337	8	ADI62190	Adi62190 Nitrilase
15	1339	75.9	337	8	ADI64311	Adi64311 Nitrilase
16	1339	75.9	337	8	ADI64301	Adi64301 Nitrilase
17	1139	64.5	333	7	ADC24025	Adc24025 Protein s
18	1139	64.5	333	8	ADH36126	Adh36126 Chemical
19	1139	64.5	333	8	ADG93827	Adg93827 Nitrilase
20	1139	64.5	333	8	ADI62424	Adi62424 Nitrilase
21	1139	64.5	333	8	ADI64545	Adi64545 Nitrilase
22	1136	64.4	332	7	ADC23873	Adc23873 Protein s
23	1136	64.4	332	8	ADH35974	Adh35974 Chemical
24	1136	64.4	332	8	ADG93675	Adg93675 Nitrilase

25	1136	64.4	332	8	ADI62272	Adi62272 Nitrilase
26	1136	64.4	332	8	ADI64393	Adi64393 Nitrilase
27	1091	61.8	355	7	ADC24039	Adc24039 Protein s
28	1091	61.8	355	8	ADH36140	Adh36140 Chemical
29	1091	61.8	355	8	ADG93841	Adg93841 Nitrilase
30	1091	61.8	355	8	ADI62438	Adi62438 Nitrilase
31	1091	61.8	355	8	ADI64559	Adi64559 Nitrilase
32	1081	61.2	345	7	ADC24067	Adc24067 Protein s
33	1081	61.2	345	8	ADH36168	Adh36168 Chemical
34	1081	61.2	345	8	ADG93868	Adg93868 Nitrilase
35	1081	61.2	345	8	ADI62466	Adi62466 Nitrilase
36	1081	61.2	345	8	ADI64587	Adi64587 Nitrilase
37	1051	59.5	334	7	ADC23739	Adc23739 Protein s
38	1051	59.5	334	8	ADH35840	Adh35840 Chemical
39	1051	59.5	334	8	ADG93541	Adg93541 Nitrilase
40	1051	59.5	334	8	ADI62138	Adi62138 Nitrilase
41	1051	59.5	334	8	ADI64259	Adi64259 Nitrilase
42	1042	59.0	353	7	ADC23945	Adc23945 Protein s
43	1042	59.0	353	8	ADH36046	Adh36046 Chemical
44	1042	59.0	353	8	ADG93747	Adg93747 Nitrilase
45	1042	59.0	353	8	ADI62344	Adi62344 Nitrilase

ALIGNMENTS

RESULT 1	
AAE05489	
ID	AAE05489 standard; protein; 337 AA.
AC	AAE05489;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Nitrilase protein #2.
XX	
KW	Nitrilase; enantiomer; alpha-substituted carboxylic acid;
XX	stereoselective; hydrolysis; amino nitrile; cyanohydrin.
OS	Unidentified.
XX	
PN	WO200148175-A2.
XX	
PD	05-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US035555.
XX	
PR	29-DEC-1999; 99US-0173609P.
XX	
PA	07-DEC-2000; 2000US-0254414P.
XX	
PI	(DIVE-) DIVERSA CORP.
XX	(MADD/) MADDEN D.
PI	Madden M, Weiner DP, Chaplin JA;
XX	
DR	WPI; 2001-465211/50.
XX	N-PSDB; AAd11167.
PT	Producing alpha-substituted carboxylic acid enantiomers by contacting
XX	aldehyde or ketone with cyanide, ammonia compound or its salt or amine
PT	and hydrolyzing the resulting amino nitrile or cyanohydrin with
XX	Nitrilase.
PS	Claim 30; Page 86-87; 87pp; English.
XX	
CC	The present invention relates to methods for producing an
CC	enantiomerically pure alpha-substituted carboxylic acid. The method
CC	involves contacting an aldehyde or ketone with a cyanide containing
CC	compound and an ammonia-containing compound or an ammonium salt or amine,
CC	and stereoselectively hydrolysing the resulting amino nitrile or
CC	cyanohydrin intermediate with a nitrilase or a polypeptide having
CC	nitrilase activity. The present sequence is a nitrilase protein #2
XX	

SQ Sequence 337 AA;
Query Match 100.0%; Score 1765; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 7.7e-188;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVACVQAPPIYMDLEATVDKTIELMEEARNNARLIAFPETWIPGYPFWLMDSP 60
DB 1 MKEAIKVACVQAPPIYMDLEATVDKTIELMEEARNNARLIAFPETWIPGYPFWLMDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
DB 61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
QY 121 TIGARRKLTKPTFEVERTLFGEGDSSLAFFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
DB 121 TIGARRKLTKPTFEVERTLFGEGDSSLAFFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
QY 181 HCAAMPSSFSLYPNAAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPSSFSLYPNAAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLLAGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
DB 241 ALLLAGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 2
ADC24119
ID ADC24119 standard; protein; 337 AA.

XX AC ADC24119;
XX DT 18-DEC-2003 (first entry)
XX DE Protein sequence (SeqID 386) exhibiting nitrilase activity.
XX KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;
XX KM enantiomer; chiral medicine.
XX OS Unidentified.
XX PN WO2003000840-A2.
XX PD 03-JAN-2003.
XX PF 15-MAY-2002; 2002WO-US015983.
XX PR 21-JUN-2001; 2001US-0300189P.
XX PR 30-JUL-2001; 2001US-0309006P.
XX PR 22-JAN-2002; 2002US-0351336P.
XX PA (DIVE-) DIVERSA CORP.
XX PA (MADD/) MADDEN D.
XX PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,
XX PI Short JM, Burk M;
XX DR WPI; 2003-201417/19.
XX DR N-PSDB; ADC24118.
XX PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-
XX PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl
XX PT lactic acid derivative and for producing pharmaceutical composition, and
XX PT food additive.
XX PS Claim 40; SEQ ID NO 386; 560pp; English.
XX CC This invention relates to nitrilases and the nucleic acids that encode

CC these enzymes thereof. Specifically, it refers to polypeptides that
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse
CC nitriles or cyanohydrins into their corresponding carboxylic acids and
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in
CC the synthesis of enantiomerically pure aromatic and aliphatic amino
CC acids, as well as hydroxy acids, which are important for the development
CC of chiral medicines. Furthermore, the present invention describes
CC nitrilases, isolated from mesophilic microorganisms, that have improved
CC activity and stability at increased pH and temperature. They are also
CC inexpensive, efficient catalysts, have broad substrate specificity and
CC are capable of chiral differentiation. This polypeptide is a protein
CC sequence that exhibits nitrilase activity of the invention.
XX
SQ Sequence 337 AA;

Query Match 99.8%; Score 1761; DB 7; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.2e-187;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVACVQAPPIYMDLEATVDKTIELMEEARNNARLIAFPETWIPGYPFWLMDSP 60
DB 1 MKEAIKVACVQAPPIYMDLKATVDKTIELMEEARNNARLIAFPETWIPGYPFWLMDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
DB 61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
QY 121 TIGARRKLTKPTFEVERTLFGEGDSSLAFFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
DB 121 TIGARRKLTKPTFEVERTLFGEGDSSLAFFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
QY 181 HCAAMPSSFSLYPNAAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPSSFSLYPNAAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

QY 241 ALLLAGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
DB 241 ALLLAGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 3
ADH36220
ID ADH36220 standard; protein; 337 AA.

XX AC ADH36220;
XX DT 11-MAR-2004 (first entry)
XX DE Chemical process monitoring-related nitrilase protein sequence SeqID386.
XX KW chemical process monitoring; biochemical process monitoring; cyanide;
XX KM high throughput system; enzyme.
XX OS Unidentified.
XX PN WO2003098187-A2.
XX PD 27-NOV-2003.
XX PF 15-MAY-2003; 2003WO-US015639.
XX PR 15-MAY-2002; 2002US-0380737P.
XX PA (DIVE-) DIVERSA CORP.
XX PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
XX PI Mcquaid J, Stege J;
XX DR WPI; 2004-142708/14.

DR N-PSDB; ADH36219.

XX Monitoring a chemical or biochemical process comprises providing a

PT reactant comprising a cyanide or a material that can be converted to

PT cyanide or a reactant that generates a cyanide or a material that can be

PT converted to cyanide.

XX

XX Claim 73; SEQ ID NO 386; 277pp; English.

PS

CC This invention relates to a novel method of monitoring chemical or

CC biochemical processes. The method involves providing a reactant

CC comprising cyanide (or a material that can be converted to a cyanide)

CC that generates as a reaction product cyanide or a material that can be

CC converted to cyanide and measuring the concentration of produced cyanide.

CC The method is useful for monitoring a chemical or biochemical process.

CC The method is effective for high throughput systems and is sufficiently

CC sensitive to detect a small amount of product. The present sequence is

CC that of a nitrilase protein which may be used in the method of the

CC invention.

XX

SQ Sequence 337 AA;

Query Match 99.8%; Score 1761; DB 8; Length 337;

Best Local Similarity 99.7%; Pred. No. 2.2e-187;

Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAPIMYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLWDSP 60

Db 1 MKEAIKVCVQAPIMYMDLKATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLWDSP 60

QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGGTLYISQWFIGDNGD 120

Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGGTLYISQWFIGDNGD 120

QY 121 TIGARRKLKPTFVERTLFEGEGSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180

Db 121 TIGARRKLKPTFVERTLFEGEGSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180

QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300

Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300

QY 301 RLIDRSPKL PVEIEGDLRPYALGKASETGAQLEEI 337

Db 301 RLIDRSPKL PVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 4

ADG93919

ID ADG93919 standard; protein; 337 AA.

XX

AC ADG93919;

XX

DT 11-MAR-2004 (first entry)

XX

DE Nitrlase enzyme amino acid sequence SeqID386.

XX

KW nitrlase; nitrile; carboxylic acid; chemical process; pH; temperature;

KW enantioselective transformation; enzyme.

XX

OS Unidentified.

XX

PN WO2003097810-A2.

XX

PD 27-NOV-2003.

XX

PF 15-MAY-2003; 2003WO-US015712.

XX

PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

XX

XX (DIVE-) DIVERSA CORP.

XX

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

XX

DR WPI; 2004-090637/09.

DR N-PSDB; ADG93918.

XX

XX

PT New isolated or recombinant nucleic acid encoding a polypeptide having

PT nitrlase activity, useful for screening enantioselective transformation.

XX

PS Claim 44; SEQ ID NO 386; 295pp; English.

XX

CC This invention is related to a novel isolated or recombinant nucleic acid

CC encoding a protein having nitrlase activity. Nitrlase's are capable of

CC converting nitrile's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated

CC nitrlase proteins of the invention have increased activity and stability

CC at increased pH and temperature when compared to those conventionally

CC used. In addition, the nucleic acid of the invention is useful for

CC screening enantioselective transformation. The present sequence is that

CC of a nitrlase enzyme of the invention.

XX

SQ Sequence 337 AA;

Query Match 99.8%; Score 1761; DB 8; length 337;

Best Local Similarity 99.7%; Pred. No. 2.2e-187;

Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAPIMYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLWDSP 60

Db 1 MKEAIKVCVQAPIMYMDLKATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLWDSP 60

QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGGTLYISQWFIGDNGD 120

Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGGTLYISQWFIGDNGD 120

QY 121 TIGARRKLKPTFVERTLFEGEGSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180

Db 121 TIGARRKLKPTFVERTLFEGEGSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180

QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300

Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300

QY 301 RLIDRSPKL PVEIEGDLRPYALGKASETGAQLEEI 337

Db 301 RLIDRSPKL PVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 5

ADI62518

ID ADI62518 standard; protein; 337 AA.

XX

AC ADI62518;

XX

DT 22-APR-2004 (first entry)

XX

DE Nitrlase polypeptide #193.

XX

KW Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;

KW (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;

KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;

KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;

KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;

KW antilipaeamic; enzyme.

XX

OS Unidentified.



XX	PN	WO2003106415-A2.
XX	PD	24-DEC-2003.
XX	PF	13-JUN-2003; 2003WO-US018840.
XX	PR	13-JUN-2002; 2002US-0389317P.
XX	PR	28-JUN-2002; 2002US-0392944P.
XX	PA	(DIVE-) DIVERSA CORP.
XX	PI	Burk M, Desantis G, Morgan B, Zhu Z;
XX	DR	WPI; 2004-090821/09.
XX	DR	N-PSDB; ADI62517.
XX	PT	Preparation of atorvastatin comprises catalytic conversion of 3-
XX	PT	hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
XX	PT	obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX	PT	acid and forming atorvastatin.
XX	PS	Claim 41; SEQ ID NO 386; 253pp; English.
XX	CC	The present invention relates to a method for preparing an atorvastatin
XX	CC	intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
XX	CC	cyano-3-hydroxybutyric acid). The method comprises optionally converting
XX	CC	epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
XX	CC	conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
XX	CC	hydroxybutyric acid with a polypeptide having nitrilase activity,
XX	CC	converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
XX	CC	acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
XX	CC	method involves whole cell processes, cell lysate process, "one pot"
XX	CC	processes, and "multi-pot" processes using a variety of parameters.
XX	CC	Atorvastatin is used, in conjunction with dietary restriction, in the
XX	CC	management of hyperlipidaemia, including hypercholesterolaemia, mixed
XX	CC	dyslipidaemia and homozygous familial hypercholesterolaemia. The present
XX	CC	sequence represents a nitrilase polypeptide obtained from an
XX	CC	environmental sample.
XX	XX	
XX	SQ	Sequence 337 AA;
	Query Match	99.8%; Score 1761; DB 8; Length 337;
	Best Local Similarity	99.7%; Pred. No. 2.2e-187;
	Matches 336; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
OY	1	MKEAIKVCVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPFWLWDSP 60
DB	1	MKEAIKVCVQAAPITYMDLKATVDKTIELMEEAARNNARLIAFPETWIPGYPFWLWDSP 60
OY	61	AWAMQFVRQYHENSLELDGPQAKRISDAKRLIGIMVTLGMSERVGTLYISQWFIGDNGD 120
DB	61	AWAMQFVRQYHENSLELDGPQAKRISDAKRLIGIMVTLGMSERVGTLYISQWFIGDNGD 120
OY	121	TIGARRKLTPTFEVERTLFGEGDSSLAVETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
DB	121	TIGARRKLTPTFEVERTLFGEGDSSLAVETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
OY	181	HCAAMPFSFLYPNAKALGPDVNVAASTRIVAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB	181	HCAAMPFSFLYPNAKALGPDVNVAASTRIVAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
OY	241	ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
DB	241	ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
OY	301	RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337
DB	301	RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337

RESULT 6  
ADI64639

ID	ADI64639	standard; protein; 337 AA.
XX	AC	ADI64639;
XX	XX	
DT	22-APR-2004	(first entry)
XX	XX	
DE	Nitrilase seq id 194.	
XX	KW	(R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
KW	KW	carboxylic acid; cyanohydrin moiety hydrolysis;
KW	KW	aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
KW	KW	chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
KW	KW	(R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
KW	KW	(R)-phenyl lactic acid derivative; & enantiomeric excess;
XX	XX	& diastomeric excess; food additive; drug intermediate; nitrilase.
OS		Unidentified.
XX	PN	US2004014195-A1.
XX	PD	22-JAN-2004.
XX	XX	
PF	15-MAY-2003; 2003US-00440523.	
XX	PR	29-DEC-1999; 99US-0173609P.
PR	07-DEC-2000; 2000US-0254414P.	
PR	28-DEC-2000; 2000US-00751299.	
PR	21-JUN-2001; 2001US-0300189P.	
PR	30-JUL-2001; 2001US-0309006P.	
PR	22-JAN-2002; 2002US-0351336P.	
PR	15-MAY-2002; 2002US-00146772.	
PR	09-SEP-2002; 2002US-00241742.	
XX	PA	(DIVE-) DIVERSA CORP.
XX	PI	Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX	DR	WPI; 2004-121569/12.
DR	DR	N-PSDB; ADI64638.
XX	PT	Novel isolated or recombinant polypeptide having nitrilase activity,
PT	PT	useful in production of food additives.
XX	PS	Claim 46; SEQ ID NO 386; 105pp; English.
XX	CC	The invention describes an isolated or recombinant polypeptide (I)
CC	CC	comprising amino acids having a sequence at least 50 % identical to a
CC	CC	sequence (S1) available in electronic form (EC) from the following web
CC	CC	site ftp.segdata.uspto.gov/sequence.html?docID=2004014195, or its
CC	CC	variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC	CC	residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC	CC	fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC	CC	hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC	CC	hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a
CC	CC	chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC	CC	producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC	CC	mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
CC	CC	derivative or an (R)-phenyl lactic acid derivative; modifying a molecule
CC	CC	and for identifying a modified compound. The inventive method is useful
CC	CC	for monitoring or determining & enantiomeric excess or & diastomeric
CC	CC	excess. (I) is useful in the production of food additives and drug
CC	CC	intermediates. This is the amino acid sequence of a nitrilase of the
CC	CC	invention.
XX	SQ	Sequence 337 AA;
	Query Match	99.8%; Score 1761; DB 8; Length 337;
	Best Local Similarity	99.7%; Pred. No. 2.2e-187;
	Matches 336; Conservative	1; Mismatches 0; Indels 0; Gaps 0;
OY	1	MKEAIKVCVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPFWLWDSP 60
DB	1	MKEAIKVCVQAAPITYMDLKATVDKTIELMEEAARNNARLIAFPETWIPGYPFWLWDSP 60



XX Claim 40; SEQ ID NO 48; 560pp; English.  
PS  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.  
XX  
SQ Sequence 337 AA;

Query Match 75.9%; Score 1339; DB 7; Length 337;  
Best Local Similarity 75.2%; Pred. No. 3.3e-140;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVCVQAAPITYMDLEATVDKTIELMEEAARNARLLIAPETWIPGYPWFLMDSP 60  
Db 1 VKKAIKVCVQAAPVFLDLDAIVDKTVALIEEAAARNGARLLIAPETWIPGYPWFLMDSP 60  
QY 61 AWAMQFVRYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120  
Db 61 AWGMQFVRYHENSLELDSPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120  
QY 121 TIGARRKLTPTFEVERTLFEGEGDSSIAVETS VGR LG LGCWEHLQPLTKYALYAQNEEI 180  
Db 121 TAGLRRLKLTPTHYERTLFEGEGDSSISTFTDPLGVLG LGCWEHLQPLSKYALYAQNEEI 180  
QY 181 HCAWPSFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HFAWPSFSIYQATEVLGPEVNVAASRIYAVEGQCFVLASCALVSPMIEMLCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILLYANLDPGVRILAKMAADPAGHYSRPDT 300  
Db 241 SLQAGGYSRIIGPDGSDLARPLGENEGILLYATLDPAAIRIYAKTAADPAGHYSRPDT 300  
QY 301 RLIDRSPLKPVVEIEGDLRPYALGKASETGAOLE 335  
Db 301 RLINRSANQPVVEVGREI-----PASAQFEVE 329

RESULT 9  
ADH35882  
ID ADH35882 standard; protein; 337 AA.  
XX  
AC ADH35882;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SegID48.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX

PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI McQuaid J, Stege J;  
XX  
DR WPI; 2004-142708/14.  
DR N-PSDB; ADH35881.

XX  
PS Claim 73; SEQ ID NO 48; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.

XX  
SQ Sequence 337 AA;  
Query Match 75.9%; Score 1339; DB 8; Length 337;  
Best Local Similarity 75.2%; Pred. No. 3.3e-140;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVCVQAAPITYMDLEATVDKTIELMEEAARNARLLIAPETWIPGYPWFLMDSP 60  
Db 1 VKKAIKVCVQAAPVFLDLDAIVDKTVALIEEAAARNGARLLIAPETWIPGYPWFLMDSP 60  
QY 61 AWAMQFVRYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120  
Db 61 AWGMQFVRYHENSLELDSPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120  
QY 121 TIGARRKLTPTFEVERTLFEGEGDSSIAVETS VGR LG LGCWEHLQPLTKYALYAQNEEI 180  
Db 121 TAGLRRLKLTPTHYERTLFEGEGDSSISTFTDPLGVLG LGCWEHLQPLSKYALYAQNEEI 180  
QY 181 HCAWPSFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HFAWPSFSIYQATEVLGPEVNVAASRIYAVEGQCFVLASCALVSPMIEMLCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILLYANLDPGVRILAKMAADPAGHYSRPDT 300  
Db 241 SLQAGGYSRIIGPDGSDLARPLGENEGILLYATLDPAAIRIYAKTAADPAGHYSRPDT 300  
QY 301 RLIDRSPLKPVVEIEGDLRPYALGKASETGAOLE 335  
Db 301 RLINRSANQPVVEVGREI-----PASAQFEVE 329

RESULT 10  
ADH35892  
ID ADH35892 standard; protein; 337 AA.  
XX  
AC ADH35892;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SegID58.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX

XX 15-MAY-2003; 2003WO-US015639.  
PF  
XX  
XX 15-MAY-2002; 2002US-0380737P.  
PR  
XX  
XX (DIVE-) DIVERSA CORP.  
PA  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI Mcquaid J, Stege J;  
XX  
DR WPI; 2004-142708/14.  
DR N-PSDB; ADH35891.  
XX  
PT Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 58; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 337 AA;  
Query Match 75.9%; Score 1339; DB 8; Length 337;  
Best Local Similarity 75.2%; Pred. No. 3.3e-140;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

KW enantioselective transformation; enzyme.  
XX  
XX OS Unidentified.  
XX  
XX PN WO2003097810-A2.  
XX  
XX PD 27-NOV-2003.  
XX  
XX PF 15-MAY-2003; 2003WO-US015712.  
XX  
XX PR 15-MAY-2002; 2002US-00146772.  
XX PR 09-SEP-2002; 2002US-00241742.  
XX  
XX PA (DIVE-) DIVERSA CORP.  
XX  
XX PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
XX DR WPI; 2004-090637/09.  
XX DR N-PSDB; ADG93582.  
XX  
XX PT New isolated or recombinant nucleic acid encoding a polypeptide having  
XX PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
XX PS Claim 44; SEQ ID NO 48; 295pp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.  
XX  
SQ Sequence 337 AA;  
Query Match 75.9%; Score 1339; DB 8; Length 337;  
Best Local Similarity 75.2%; Pred. No. 3.3e-140;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;



```

XX XX Nitriylase enzyme amino acid sequence Segid58.
XX KM nitriylase; nitriyl; carboxylic acid; chemical process; pH; temperature;
KM KM enantioselective transformation; enzyme.
XX OS Unidentified.
XX PN WO2003097810-A2.
XX PD 27-NOV-2003.
XX PF 15-MAY-2003; 2003WO-US015712.
XX PR 15-MAY-2002; 2002US-00146772.
XX PR 09-SEP-2002; 2002US-00241742.
XX PA (DIVE-) DIVERSA CORP.
XX PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;
XX PI WPI; 2004-090637/09.
XX DR N-PSDB; ADG93592.
XX PT New isolated or recombinant nucleic acid encoding a polypeptide having
XX PT nitriylase activity, useful for screening enantioselective transformation.
XX PS Claim 44; SEQ ID NO 58; 295pp; English.
XX CC This invention is related to a novel isolated or recombinant nucleic acid
XX CC encoding a protein having nitriylase activity. Nitriylase's are capable of
XX CC converting nitriyl's directly to carboxylic acids and have great
XX CC potential for use in industrial chemical processes. The isolated
XX CC nitriylase proteins of the invention have increased activity and stability
XX CC at increased pH and temperature when compared to those conventionally
XX CC used. In addition, the nucleic acid of the invention is useful for
XX CC screening enantioselective transformation. The present sequence is that
XX CC of a nitriylase enzyme of the invention.
XX SQ Sequence 337 AA;

Query Match 75.9%; Score 1339; DB 8; Length 337;
Best Local Similarity 75.2%; Pred. No. 3.3e-140;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1,

QY 1 MKEAIKVCVQAAPITMDLEATVDKTEIMEBAARNRRLIAFPETWIPGYPWLMDSP 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 VKEAIKVCVQAAPVFLDLDTATVDKTVALLIEAARNGARLIAFPETWIPGYPWLMDSP 60

QY 61 AWAMQFVRYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 AWAMQFVRRYHENSLELDSPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120

QY 121 TIGARRKLPFTFVERTLFGEGDSSLAFFSVGRLLGGLCCWEHLQPLTKYALYAONEBI 180
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 TAGLRKRLKPTHTVERTLFGEGDSSLSFTFPLGVLLGLCCWEHLQPLSKYALYAONEBI 180

QY 181 HCAAMPFSFUYENAAKALGPDVNVAAASRIYAVEGQCFFVLASCALVSGSMIDMLCTDDEKH 240
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 HCAAMPFSFUYENAAKALGPDVNVAAASRIYAVEGQCFFVLASCALVSGSMIDMLCTDDEKH 240

QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGLYANLDPGVRILAKMADDPAGHSRDPIT 300
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 SLLAGGHSRIIGPDGDLVAPLAENEGLYANLDPGVRILAKMADDPAGHSRDPIT 300

QY 301 RLLIDRSPKLPVEIEGDLRPYALGKASGTGQLE 335
||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 301 RLLIDRSPKLPVEIEGDLRPYALGKASGTGQLE 335

AD162180 standard; protein; 337 AA.

```

XX		AD162180;		
AC		(first entry)		
XX	D1	Nitrilase polypeptide #24.		
DE		Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate; (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin; 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile; 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid; mixed hyperlipidaemia; homozygous familial hypercholesterolaemia; antihypertensive enzyme.  Unidentified.		
OS		WO2003106415-A2.		
PN		24-DEC-2003.		
PD		13-JUN-2003; 2003WO-US018840.		
PF		13-JUN-2002; 2002US-0389317P.		
PX		PR 28-JUN-2002; 2002US-0392944P.		
PA		(DIVE-) DIVERSA CORP.		
XI	Burk M,	Desantis G, Morgan B, Zhu Z;		
PI		WPI; 2004-090821/09.		
DR		N-PADB; AD162179.		
PT		Preparation of atorvastatin comprises catalytic conversion of 3-		
PT		hydroxyglutaronitrile by polypeptide with nitrilase activity, converting obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric acid and forming atorvastatin.		
PS	Claim 41;	SEQ ID NO 48; 253bp; English.		
XX	The present invention relates to a method for preparing an atorvastatin intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-cyano-3-hydroxybutyric acid). The method comprises optionally converting epichlorohydrin or equivalent to 3-hydroxygluteronitrile, catalytic conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-hydroxybutyric acid with a polypeptide having nitrilase activity, converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The method involves whole cell processes, cell lysate process, "one pot" processes, and "multi-pot" processes using a variety of parameters. Atorvastatin is used, in conjunction with dietary restriction, in the management of hyperlipidaemia, including hypercholesterolaemia, mixed dyslipidaemia and homozygous familial hypercholesterolaemia. The present sequence represents a nitrilase polypeptide obtained from an environmental sample.			
SQ		Sequence 337 AA:		
	Query Match	75.9%; Score 1339; DB 8; Length 337;		
	Best Local Similarity	75.2% Pred. No. 3.3e-140;		
Matches	Conservative 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;			
OY	1 MKKAIKYAVCQAAPITMDLEATVDTKTIELMEBAARNNARLIAPEETWIPGYPMFLWDSP	60		
DB	1 VKKAIKYVCQAAPVFLLDLDAVTUXTVALIEBAAANGARLIAPETWIPGYPMFLWDSP	60		
OY	61 AWAMQFVRQYHNSLSLDGPQAKRTSDAKRLGIMWTLMGSERVGTLYISQMFIGDNGD	120		
DB	61 AMWMGFVRQRYSNLVLDSPPAQKRISSEAAGRAIGTVALSGSERVSGLTYMGMWLIDDKGE	120		
OY	121 TIGARRKLKFTFVERTLFEGEDSSLAIVFETS VGR LG LCCMEHQPLTKVALLYAQNNEI	180		
DB	121 TAGLRKKLPHTHVERTLFEGEDSSISTFDTPPLGVLG LCCMEHI QPLSKYALYAQNNEI	180		

[illegible]

Query Match	75.9%;	Score 1339;	DB 8;	Length 337;
Best Local Similarity	75.2%;	Pred. No. 3.3e-140;		
Matches 252;	Conservative 30;	Mismatches 47;	Indels 6;	Gaps 1;
QY	1	MKEAIKVCVQAAPITYMDLEATVDKTIEMEAAARNARLIAFPETWIPGYPWFMLDLP	60	
Db	1	VKEAIKVCVQAAPVFLDLDATVDKTVALLIEAARNGARLIAFPETWIPGYPWFMLDLP	60	
QY	61	AMAMQFVRQYHENSLELDGPOAKRISDAKRLIGIMVTLGMSERVGGTLTYSQWFIGDNGD	120	
Db	61	AMGMQFVRQYHENSLELDSPQAKRISDAQRAGIYVALGYSERVSGILYMGQMLIDDKGE	120	
QY	121	TIGARRKLKPTFEVERTLFGEGDSSLAVEFETSVGRLGGLCWEHLQPLTKYALYAQNEEI	180	
Db	121	TAGLRRLKLKPTFEVERTLFGEGDSSLSFTDPLGVLGGLCWEHLQPLSKYALYAQNEEI	180	
QY	181	HCAAMPSSFLYPNAKALGPDVNVAASTRIVAVEGQCFVLASCALVSQSMIDMLCTDDEKH	240	
Db	181	HFAAMPSSFSTRQATEVLGPVNVAASTRIVAVEGQCFVLASCALVSPENIEMLCTDESKH	240	
QY	241	ALLLAGGHSRIITGPDGDLVAPLPAENEEGILYANLDPGVRIIAKMAADPAGHYSPDIT	300	
Db	241	SLLQAGGGYSRIITGPDGSDLARPLGENEEGILYATLDPARIYAKTAADPAGHYSPDVT	300	
QY	301	RLIIDRSPKLPVEIEIGDLRPYALGKASSETGAQLE	335	
Db	301	RLIINRSANQPVVEVGREI-----PASAQGFEVE	329	



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:00:30 ; Search time 33.552 Seconds  
(without alignments)  
830.404 Million cell updates/sec

Title: US-09-751-299-4

Perfect score: 1765  
Sequence: 1 MKEAIKVCQAQAPIYMDL.....DLRPVALGKASGTGAQLEEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	955.5	54.1	344	1	US-08-690-493-1	Sequence 1, Appli
2	822	46.6	356	2	US-09-806-876A-2	Sequence 2, Appli
3	776	44.0	369	2	US-09-823-373-5	Sequence 5, Appli
4	776	44.0	369	2	US-09-823-373-14	Sequence 14, Appli
5	638	36.1	354	1	US-08-447-702-5	Sequence 5, Appli
6	638	36.1	354	1	US-08-465-615-5	Sequence 5, Appli
7	152	8.6	267	2	US-09-902-540-14601	Sequence 14601, A
8	139.5	7.9	315	2	US-10-105-294B-2	Sequence 2, Appli
9	139	7.9	311	1	US-07-917-111-5	Sequence 5, Appli
10	139	7.9	311	1	US-07-917-111-6	Sequence 6, Appli
11	139	7.9	311	1	US-08-479-638-5	Sequence 5, Appli
12	139	7.9	311	1	US-08-479-638-6	Sequence 6, Appli
13	139	7.9	311	1	US-08-294-871A-70	Sequence 70, Appli
14	139	7.9	311	2	US-08-876-398A-70	Sequence 70, Appli
15	133	7.5	303	1	US-08-294-871A-46	Sequence 46, Appli
16	133	7.5	303	2	US-08-876-398A-46	Sequence 46, Appli
17	132	7.5	303	1	US-08-294-871A-36	Sequence 36, Appli
18	132	7.5	303	2	US-08-876-398A-36	Sequence 36, Appli
19	131	7.4	303	1	US-08-294-871A-4	Sequence 4, Appli
20	131	7.4	303	1	US-08-294-871A-16	Sequence 16, Appli
21	131	7.4	303	1	US-08-294-871A-34	Sequence 34, Appli
22	131	7.4	303	1	US-08-294-871A-48	Sequence 48, Appli
23	131	7.4	303	1	US-08-294-871A-58	Sequence 58, Appli
24	131	7.4	303	1	US-08-294-871A-68	Sequence 68, Appli
25	131	7.4	303	2	US-08-876-398A-4	Sequence 4, Appli
26	131	7.4	303	2	US-08-876-398A-16	Sequence 16, Appli
27	131	7.4	303	2	US-08-876-398A-34	Sequence 34, Appli

28	131	7.4	303	2	US-08-876-398A-48	Sequence 48, Appli
29	131	7.4	303	2	US-08-876-398A-58	Sequence 58, Appli
30	131	7.4	303	2	US-08-876-398A-68	Sequence 68, Appli
31	130	7.4	303	1	US-08-294-871A-38	Sequence 38, Appli
32	130	7.4	303	1	US-08-294-871A-44	Sequence 44, Appli
33	130	7.4	303	1	US-08-294-871A-50	Sequence 50, Appli
34	130	7.4	303	1	US-08-294-871A-56	Sequence 56, Appli
35	130	7.4	303	1	US-08-294-871A-60	Sequence 60, Appli
36	130	7.4	303	2	US-08-876-398A-38	Sequence 38, Appli
37	130	7.4	303	2	US-08-876-398A-44	Sequence 44, Appli
38	130	7.4	303	2	US-08-876-398A-50	Sequence 50, Appli
39	130	7.4	303	2	US-08-876-398A-56	Sequence 56, Appli
40	130	7.4	303	2	US-08-876-398A-60	Sequence 60, Appli
41	130	7.4	304	1	US-08-900-711-5	Sequence 5, Appli
42	130	7.4	304	1	US-08-815-356-2	Sequence 2, Appli
43	130	7.4	304	1	US-08-415-343B-5	Sequence 5, Appli
44	130	7.4	304	1	US-08-762-433-1	Sequence 1, Appli
45	130	7.4	304	1	US-09-001-219-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-1  
; Sequence 1, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000e1 Nitrlilase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steindberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acid residues  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Gordona terrae  
; STRAIN: MA-1  
; CELL TYPE: unicellular organism  
; FEATURE:  
; OTHER INFORMATION: Xaa is Met or a deletion  
US-08-690-493-1  
Query Match 54.1%; Score 955.5; DB 1; Length 344;  
Best Local Similarity 58.3%; Pred. No. 2.4e-97;  
Matches 182; Conservative 47; Mismatches 82; Indels 1; Gaps 1;





```
; CURRENT APPLICATION NUMBER: US/09/823,373
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Acidovorax facilis
US-09-823-373-14

Query Match      44.0%; Score 776; DB 2; Length 369;
Best Local Similarity 47.6%; Pred. No. 2.4e-77;
Matches 150; Conservative 61; Mismatches 100; Indels 4; Gaps 3;

QY 8 ACVQAAPLYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPWLMDSPAWAMQFV 67
DB 11 ATVQAEPWLDADATIDKISIGIIEEAQKASLIAFPEVFIPGYPWAWLGDVKYSLSFT 70

QY 68 RQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGTLTYSQWFIGDNGDTIGARRK 127
DB 71 SRYHENSLELDGDRMRRLQLAARRNKIALVMGYSEREAGSRVLSQVFI DERGEIVANRRK 130

QY 128 LKPTFVERTLFGEGDSSSLAVFETSVGRLGCLCWEHLQPLTKYALYAQNEIHCAMP 187
DB 131 LKPTFVERTIYGEENGDTFLTHDFAFGVGLNCWEHFQPLSKFMYTSLGEQVHVASWPA 190

QY 188 FS-LYPNAKALGPDVNVAASRIYAVEGQFVLASCALVSQSMIDMLCTDDEKHALLAG 246
DB 191 MSPLOPDVFO-LISEANATVTRSYAIEGQTFVLCSTQVIGPSAIEFTFCINDEQALLPQG 249

QY 247 GGHRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRDPDITRLIDR 306
DB 250 CGWARIYGDGSELAKPLAEDAEGILYAEIDLQILLAKAGADPVGHYSRPDVLVSQFDP 309

QY 307 SPKLPV--VEIEGDL 319
DB 310 RNHTPVHRIGIDGRL 324

RESULT 5
US-08-447-702-5
; Sequence 5, Application US/08447702
; Patent No. 5629190
; GENERAL INFORMATION:
; APPLICANT: Petre, Dominique
; APPLICANT: Cerbeleaud, Edith
; APPLICANT: Levy-Schill, Sophie
; APPLICANT: Crouzet, Joel
; TITLE OF INVENTION: Polypeptides Possessing A Nitrilase
; TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,
; Patent No. 5629190
; TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them
; TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy-
; TITLE OF INVENTION: lates by Means of Said Polypeptides
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,702
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 530
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/194,588
; FILING DATE: 10-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92-09-882
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 003025-019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 354 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-447-702-5

Query Match      36.1%; Score 638; DB 1; Length 354;
Best Local Similarity 41.2%; Pred. No. 5e-62;
Matches 134; Conservative 62; Mismatches 107; Indels 22; Gaps 5;

QY 5 IKVACVQAAPLYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPWLMDSPAWAMQFV 59
DB 7 VKVAVQAAPVFMNLLEATVDKTKLIAEASMGAKVIGFPAFIPGYPIWITSNMDPTG 66

QY 60 PAMWQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGTLTYSQWFIGDNG 119
DB 67 MNAVLF-----KNAIEIPSEYQQLSDAKKNGVYCVSVSEKDNASLYLTQLWFDPN 121

QY 120 DTIGARRKLTFTFVERTLFGEGDSSSLAVFETSVGRLGCLCWEHLQPLTKYALYAQNEE 179
DB 122 NLIGKRRKFTPTSSERAVWGDGSMAPVFKTEYGNLGLQCWEHALPLNIAAMGSLNEQ 181

QY 180 IHCAAMPFSFLYPNAKALGPDVNVA-----SRIYAVEGQFVLASCALVSQSMI 230
DB 182 VHVASWPAF--VPKGAVSSRVSSVCASTNAMHQIISQFYAISNQVYVIMSTNLVGQDMI 239

QY 231 DMLCTDEKHALLAGGGHRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADP 290
DB 240 DMIGKDEFSKNFLPLGSGNTAIIIS-NTGEILASIPQDAEGIAVAEIDLNQIIYGKWLDP 298

QY 291 AGHSRDPDITRLIDRSPKLPVEI 315
DB 299 AGHYSTPGFLSTFDQSEHYVVKKI 323

RESULT 6
US-08-465-615-5
; Sequence 5, Application US/08465615
; Patent No. 5635391
; GENERAL INFORMATION:
; APPLICANT: PETRE, Dominique
; APPLICANT: CERBELEAUD, Edith
; APPLICANT: LEVY-SCHILL, Sophie
; APPLICANT: CROUZET, Joel
; TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE
; TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,
; TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM
; TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO
; TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/465,615
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/194,588
;     FILING DATE: 10-FEB-1994
;     APPLICATION NUMBER: FR 9209882
;     FILING DATE: 10-AUG-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Crane-Feury, Sharon E
;     REGISTRATION NUMBER: 36,113
;     REFERENCE/DOCKET NUMBER: 003025-015
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703) 836-6620
;     TELEFAX: (703) 836-2021
;   INFORMATION FOR SEQ ID NO: 5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 354 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;     MOLECULE TYPE: protein
;   US-08-465-615-5

```

Query Match 36.1%; Score 638; DB 1; Length 354;  
Best Local Similarity 41.2%; Pred. No. 5e-62;  
Matches 134; Conservative 62; Mismatches 107; Indels 22; Gaps 5;

```

QY 5 IKVACVQAAPYIMDLAATVDKTIELMEEAARNNARLIAPEETWIPGYPFWLWDS----- 59
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 7 VKVAAVQAAPVFMNLAEATVDKCKLIAEAASMGAKVIGFEAFIPGYPYWITSNMDFTG 66
   :||:|-----KNAIEIPSKVQIISDAAKNGVYVCSYSEKDNASLYLTQLWEDPNG 121
QY 60 PAWAMQFVRQYHENSLELDGPOAKRISDAAKRLGIMWTLGMSERVGGLYISQWFIGDNG 119
   ||:|-----:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 67 MMWAVLF-----KNAIEIPSKVQIISDAAKNGVYVCSYSEKDNASLYLTQLWEDPNG 121
QY 120 DTIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEE 179
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 122 NLIGKHKFKPTSSERAVWGDGDSMAVFKTEYGNLGLQCEHALPLNIAAMGSLNEQ 181
QY 180 IHCAWPSFSLYPNAKALGPDVNVAA-----SRIYAVEGQCFLVASCALVSQSMI 230
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 182 VHVASWPAF--VPKGAVSSRVSSVCASTNAMHQIISQFYAISNQVYVIMSTNLVGQDMI 239
QY 231 DMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADP 290
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 240 DMIGKDEFKSNFLPLGSGNTAIIIS-NTGEILASIPQDAEGIAVAEIDLNQIITYGKWLLDP 298
QY 291 AGHYSRPDITRLLIDRSFKLPVVEI 315
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 299 AGHYSTPGFLSLTFDQSEHVPVKTI 323

```

RESULT 7  
US-09-902-540-14601  
; Sequence 14601, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825

```

; SEQ ID NO 14601
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
; US-09-902-540-14601

```

Query Match 8.6%; Score 152; DB 2; Length 267;  
Best Local Similarity 24.2%; Pred. No. 3.4e-08;  
Matches 67; Conservative 36; Mismatches 116; Indels 58; Gaps 10;

```

QY 21 ATVDK-----TIELMEEAARNNARLIAPEETWIPGYPFWLWDS PANAMQFVRQYHENS 74
   :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 11 STADKAHNLEAATRLVRAVALGARLVGLPEN-----FSWMGPEP-----ERQDAA 56
QY 75 LELDGPQAKRISDAKRLGIMWTLGM---SERVGTLYISQWFIGDNGDTIGARRKCLKPT 131
   ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 57 EGLDGPTLSQMASLARELKVTLLAGSVLETGAPGRLYNTSVLFGPGGERLAYYRKIHLE 116
QY 132 FVERTLFGEG-----DGSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEIH 181
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 117 DVE---VGDAITYGESAAVAPGTEVVSAAETEVGRIGLSVCYDLRFP-ELYYRLSRGATL 172
QY 182 CAWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFLVASCALVSQSMIDMLCTDDEKHA 241
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 173 LAVPAFTL-----MTGKDHEVLLRARAIEHQAYVLAP-----AQGGRRHS 213
QY 242 LLLAGGHSRIIGPDGDLVAPLAENEGILYANLDP 278
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 214 ANRVTYGHALVDPWG--LVTARASEGEGALAPVDP 248

```

RESULT 8  
US-10-105-294B-2  
; Sequence 2, Application US/10105294B  
; Patent No. 6800464  
; GENERAL INFORMATION:  
; APPLICANT: DRAUZ, KARLHEINZ  
; APPLICANT: MAY, OLIVER  
; APPLICANT: BOMMARIUS, ANDREAS  
; APPLICANT: SYLDATK, CHRISTOPH  
; APPLICANT: ALTENBUCHNER, JOSEF  
; APPLICANT: WERNER, MARKUS  
; APPLICANT: SIEMANN-HERZBERG, MARTIN  
; TITLE OF INVENTION: D-Carbamoylase from Arthrobacter crystallopoietes  
; FILE REFERENCE: 220024US0X  
; CURRENT APPLICATION NUMBER: US/10/105,294B  
; CURRENT FILING DATE: 2000-08-11  
; PRIOR APPLICATION NUMBER: DE 101 14 999.9  
; PRIOR FILING DATE: 2001-03-26  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 2  
; LENGTH: 315  
; TYPE: PRT  
; ORGANISM: ARTIFICIAL SEQUENCE  
; FEATURE:  
; OTHER INFORMATION: SYNTHETIC DNA  
; US-10-105-294B-2

Query Match 7.9%; Score 139.5; DB 2; Length 315;  
Best Local Similarity 24.3%; Pred. No. 1.1e-06;  
Matches 77; Conservative 47; Mismatches 134; Indels 59; Gaps 16;

```

QY 23 VDKTIELMEEAARNNARLIAPE---TWIPGYPFWLWDS PANAMQFVRQYHENSLELD 78
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 25 VARLIALLEBAASQGAELVFPDELTLTTFP-RTWFEEDG-----FEEYFDKSMEND 75
QY 79 GPQAKRISDAKRLGIMWTLGMSERVGG-TLYISQWFIGDNGDTIGARRK----- 127
   :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 76 --DVAPLPERAKDLGVGYALELTSDEKRYNTSILVNKHGDIVGKYRKMLPGHADNR 133
QY 128 --LKPTFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEIHCAAW 185
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:

```

Db 134 EGLPNQHLKKYFREGD-LGFGVFDFHGVQVMCLCNDRRWDEVYRSLALQGAELVVLGY 192  
QY 186 -----PSFSLYPNAKALGPDVNVVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEK 239  
Db 193 NTPDFVFWQOEHPHAKMF-----THLISLQAGAYQNSVFVAAA-----GKSGF-----EDGH 239  
QY 240 HALLLAGGHSRIIGPDGDLVAPLAENEGILY-ANLDPGVRILAKMAADPAGHYSRPD 298  
Db 240 HMI-----GGSVAAPSGEILAKAAGEDEVVVKADIDMG-KPKYESVFDEFAAH-RRPD 292  
QY 299 ITRLLIDRSPK--LPV 312  
Db 293 AYGIIAERKRGAPLPV 309

## RESULT 9

US-07-917-111-5  
; Sequence 5, Application US/07917111  
; Patent No. 5565344  
; GENERAL INFORMATION:  
; APPLICANT: Namba, Hirokazu  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Yajima, Kazuyoshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
; TITLE OF INVENTION: ACIDS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/917,111  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Player Esq., William E.  
; REGISTRATION NUMBER: 31,409  
; REFERENCE/DOCKET NUMBER: P-500-23486  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-887-0400  
; TELEFAX: 202-835-0605  
; TELEX: 440706  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-917-111-5

Query Match 7.9%; Score 139; DB 1; Length 311;

Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAPAPIYMD--LEATVDKTIEMEBARNNARLIAFPETWIPGYPMFLWDSPAW 62  
Db 4 VNAAAQMGPIISRSETRKDTVRRLIALMREAKARGSDLVFTELALTTF-----FPRW 56  
QY 63 AMQFVRQYHENSLEID-----GPOAKRISDAKRLGIMVTIGMSE--RVGG--TLYI 110  
Db 57 VI-----EDEAEIDSFYEKEMPGPETQPLFDEAKRLEIGFYLYAEIAEBEGKRKRFN 109  
QY 111 SQWFIGNDGTIGARRKL-----KPTFVERTLFGBGDSSLAVFETSVGRLG 158  
Db 110 TSLVDKRGRIYGVKRVHLPGHKEPQPGRKQHLKKRYFERGD-LGFGVRAAFDGVGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEIHCAM--PSFSLYPNAKALGPDVNVVAASRIYAVEGQC 216  
Db 169 CICNDRWPETRYVMGLQGVEMVMLGYNTPYHTGHDDIDSLTFQHNHLSMQAGAYQNST 228  
QY 217 FVLASCALVSQSMIDMLCTDDEKHALLAGGHSRIIGPDGDLVAPLAENEGILYA-- 274  
Db 229 WVIQTAK-----CGTEGSKMV---GQSVIVAP-SGEIVAMACTIEDIITARC 273  
QY 275 NLDPGVRILAKMAADPAGHYSRPDITRLLIDRSPKLP 311  
Db 274 DLDMGKR-YRETFDFARH-REPDAYRLIVERKGAVP 308

## RESULT 10

US-07-917-111-6  
; Sequence 6, Application US/07917111  
; Patent No. 5565344  
; GENERAL INFORMATION:  
; APPLICANT: Namba, Hirokazu  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Yajima, Kazuyoshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
; TITLE OF INVENTION: ACIDS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/917,111  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Player Esq., William E.  
; REGISTRATION NUMBER: 31,409  
; REFERENCE/DOCKET NUMBER: P-500-23486



```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-0400
TELEFAX: 202-835-0605
TELEX: 440706
INFORMATION SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Pseudomonas
STRAIN: KNK 003A (FERM BP-3181)

```

Query Match	7.9%;	Score 139;	DB 1;	Length 311;
Best Local Similarity	23.7%;	Pred. No. 1.2e-06;		
Matches	80;	Conservative 49;	Mismatches 146;	Indels 62;
				Gaps 15;

```
OY      5 IKVACVOAFIYMD--LEATVDKTELMEEAARNNARLIAFPETWPGYPFWLWDSPAW 62
         : | | | | : | | | | : | | | | : | | | |
Db      4 VNAAAAQMGPIRSSETRKDIVRRLLIALMBEAKARGSDLVFTELALTTF-----FPRW 56
OY      63 AMQFVRQYHENSLELD-----GPOAKRISDAKRIGIMVTLCMSE--RYGG--TLYI 110
         : | | | | | | | | | | | | | | | | | |
Db      57 VI-----EDEAEILDSFEYEKEMPGETQLPFDCAKRLIEGYLYIAELAECCGRKKRFN 109
OY      111 SQWFIGNDGTIGARRKL-----KPTEVERTLFEGEGSSLAVFETSVGRLGG 158
        : : : : : | | | | : | | | | : | | | |
Db      110 TSLIDRSGRIVGKYRKVHLPGHKEPQGRKHQHLEKRYFERPD-LGFYWRAFDGVGMG 168
OY      159 LCCWEHLQPLTKYALYAONEIHCAAW--PSFSLYPNAKALGPDVNVAAARIYAVEGC 216
        : | | | | : | | | | : | | | | : | | | |
Db      169 CINCNDRWPEYYRMGLQGVEVMVLGNTPPYDHGDIDDLSLTQFHNLHSMQAGAYONST 228
OY      217 FYLASCALVSQSMDMLCTDDEKHALLLAGGHSRIIPDGDDLVAFLAENEEGILIYA-- 274
        : : : : : | | | | : | | | | : | | | |
Db      229 WVI GTAK-----CGTEESGSKMV----GQSVIVAP-SGEIVAMACTIEDIEITTARC 273
OY      275 NUDPGVRILAKMAADPA GHYSRPDITRLLIDRSPKLP 311
        : | | | | : | | | | | | | | : |
Db      274 DLDMGKR-YRETIFDFARH-REPDAYRLIVERKA V P 308
```

RESULT 11  
 US-08-479-638-5  
 ; Sequence 5, Application US/08479638  
 ; Patent No. 5695968  
 ;  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Namba, Hirokazu  
 ; APPLICANT: Yamada, Yukio  
 ; APPLICANT: Takano, Masayuki  
 ; APPLICANT: Ikenaka, Yasuhiro  
 ; APPLICANT: Takahashi, Satomi  
 ; APPLICANT: Yajima, Kazuyoshi  
 ; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
 ; TITLE OF INVENTION: ACIDS  
 ;  
 ; NUMBER OF SEQUENCES: 6  
 ;  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Wegner, Cantor, Mueller & Player  
 ; STREET: 1233 20th Street, N.W., Suite 300  
 ; CITY: Washington  
 ; STATE: D.C.  
 ;  
 ; ZIP: 20036-8218  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US//08/479,638  
 ; FILING DATE: 07-JUN-1995

```

1 CLASSIFICATION: 435
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER: US 07/917,111
4 FILING DATE: 07-AUG-1992
5 APPLICATION NUMBER: JP 400848/1990
6 FILING DATE: 07-DEC-1990
7 PRIOR APPLICATION DATA:
8 APPLICATION NUMBER: JP 407922/1990
9 FILING DATE: 27-DEC-1990
10 PRIOR APPLICATION DATA:
11 APPLICATION NUMBER: JP 078840/1991
12 FILING DATE: 11-APR-1991
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: PCT/JP91/01696
15 FILING DATE: 06-DEC-1991
16 ATTORNEY/AGENT INFORMATION:
17 NAME: Player Esq., William E.
18 REGISTRATION NUMBER: 31,409
19 REFERENCE/DOCKET NUMBER: P-500-23486
20 TELECOMMUNICATION INFORMATION:
21 TELEPHONE: 202-887-0400
22 TELEFAX: 202-835-0605
23
24 TELEX: 440706
25 INFORMATION FOR SEQ ID NO: 5:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 311 amino acids
28 TYPE: amino acid
29 TOPOLOGY: linear
30 MOLECULE TYPE: protein
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
```

Query Match	7.9%;	Score 139;	DB 1;	Length 311;
Best Local Similarity	23.7%;	Pred. No. 1.2e-06;		
Matches	80;	Conservative	49;	Mismatches 146;
				Indels 62;
				Gaps 15;

QY		5 IKVACVOAPIYMD--LEATVDKTIELMEEAARNARLIAFPETWIPGYFWFLMDSPAW	62
Dd		4 VNAAAQMGPIRSSETRKDIVRLIALMREAKARGSDLVFTELALTTF-----FPRW	56
QY		63 AMQFVRQYHENSLELD-----GPOAKRISDAKRIGIMVLGMSE--RVGG--TLVI	110
Dd		57 VI-----EDEAELDSFEKEMPGPETOPLFDEAKRLEIGFYLYAEALAEEGKKRRFN	109
QY		111 SQWFIGNGDITIGARKL-----KPTFVERTLFGEKGSSLAVFETSVGRLGG	158
Dd		110 TSI LVBSGRIVGKYRKVHLPGHKERQPGRKHQHLERYFERPD-LGFGWRAFDGVGM	168
QY		159 LCCWEHIQPLTKYALYAQNIEIHCAW--PSFSLYPNAKALGPDVNVAAARIYAVEGQC	216
Dd		169 CICI NDRMPETYRYVMGLQGVEWMVLGYNTPYDHTGHDDISLTQHFNHLSMQAGAYONST	228
QY		217 FVLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGILVA--	274
Dd		229 WVIGTAK-----CGTEEGSKMV---GQSIVIAP-SGEIVAMACTIEDEIIYARC	273
QY		275 NLDPGVRIIAKMAADPA GHYSRPDI TRLLIDRS PKLP	311
Dd		274 DLD MGKR-YRETIFDFARH-REPDA YRLIVERKGA VP	308

RESULT 12  
 US-08-479-638-6  
 ; Sequence 6, Application US/08479638  
 ; Patent No. 5695968  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Namba, Hirokazu  
 ; APPLICANT: Yamada, Yukio  
 ; APPLICANT: Takano, Masayuki  
 ; APPLICANT: Ikenaka, Yasuhiro  
 ; APPLICANT: Takahashi, Satomi  
 ; APPLICANT: Yajima, Kazuyoshi  
 ; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
 ; TITLE OF INVENTION: ACIDS  
 ;

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wegner, Cantor, Mueller & Player  
STREET: 1233 20th Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20036-8218  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,638  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player Esq., William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P-500-23486  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-0400  
TELEFAX: 202-835-0605  
TELEX: 440706  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Pseudomonas  
STRAIN: KNK 003A (FERM BP-3181)  
US-08-479-638-6

Query Match 7.9%; Score 139; DB 1; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;  
QY 5 IKVACVQAPIYMD--LEATVDKTELMEEAARNNARLIAFPETWIPGYPMFLWLDSPAW 62  
DB 4 VNAAAQMGFISRSSETRKDTVRRLIALMREAKARGSIDLVPFTELALTF-----FPRW 56  
QY 63 AMQFYROYHENSLELD-----GFOAKRISDAKRLGIMVTIGMSE--RVGG--TLYI 110  
DB 57 VI-----EDEABLDSPYEKEMPGPETQPLFDEAKRLIEGYIGYAELEAEEGGRKRFRN 109  
QY 111 SQWFIGDNGDTIGARRKL-----KPTFVERTLFGEGDSSSLAVFETSVGRLG 158  
DB 110 TSLVDRSGRIVGKYRKVHLPGHKEPQGRKHQLLEKRYFEPGD-LGFGVWRAFDGVMGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEEIHCAAW--PSFSLYPNAKALGPDVNVAAASRIYAVEGQC 216  
DB 169 CICNDRWPEFYRVMLQGVEMVMLGNTPYDHTGDDIDSLQFHNHLSMQAGAYQNST 228  
QY 217 FYVLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGGLVAPLAENEGLYA-- 274  
DB 229 WVIQTAK-----CGTEEGSKV-----GQSVIVAP-SGEIVAMACTIEDEITITARC 273

QY 275 NUDPGVRILAKMADPAGHSRPDITRLIDRSPKLP 311  
DB 274 LIDMGR-KRETIFDPAH-REPDAYRLIVERKGAVP 308  
RESULT 13  
US-08-294-871A-70  
Sequence 70, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko  
APPLICANT: Hiraishi, Yoshiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-70

Query Match 7.9%; Score 139; DB 1; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAAPRYMD--LEATVDKTIELMEBAARNNARLIAFPETWIPGYPWFLWLDSPAW 62  
DB 4 VNAAAQMGPIRSSETRKDTVRRLIALMREAKARGSDLVFTLALTTF-----FPRW 56  
QY 63 AMQFVROYHENSLELD-----GPOAKRISDAKRLGIMVTLGMSR--RVGG--TLYI 110  
DB 57 VI-----EDEALDSFYEKEMPGPETQPLFDEAKRLGIFLYGYALAELEGGRKRFRN 109  
QY 111 SQWFIGNGDTIGARRKL-----KPTFVERTLFGEGDSSLAVFETSVGRLLG 158  
DB 110 TSLVDRSGRIVGKYRKVHLPGHKEPQGRKHQHLEKRYFEPGD-LGFGVWRAFDGVMGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEEIHCAAW--PSFLYPNAKALGPDVNVAAASRIYAVEGQC 216  
DB 169 CICNDRWPETRYVMGLQGVEMVMLGYNTPYDHTGHDIDSLTQFHNHLSMQAGAYQNST 228  
QY 217 FVLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYA-- 274  
DB 229 WVIQTAK-----CGTEEGSKMV---GQSVIVAP-SGEIVAMACTIEDEIITARC 273  
QY 275 NLDPGVRIAKMAADPAGHYSRPDITRLLIDRSPKLP 311  
DB 274 DLDMGKR-YRETIFFDFAH-REPDAYRLIVERKGAVP 308

RESULT 14  
US-08-876-398A-70  
Sequence 70, Application US/08876398A  
Patent No. 6083752  
GENERAL INFORMATION:

APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NANBA, Hirokazu  
APPLICANT: TAKANO, Masayuki  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-70

Query Match 7.9%; Score 139; DB 2; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAAPRYMD--LEATVDKTIELMEBAARNNARLIAFPETWIPGYPWFLWLDSPAW 62  
DB 4 VNAAAQMGPIRSSETRKDTVRRLIALMREAKARGSDLVFTLALTTF-----FPRW 56  
QY 63 AMQFVROYHENSLELD-----GPOAKRISDAKRLGIMVTLGMSR--RVGG--TLYI 110  
DB 57 VI-----EDEALDSFYEKEMPGPETQPLFDEAKRLGIFLYGYALAELEGGRKRFRN 109  
QY 111 SQWFIGNGDTIGARRKL-----KPTFVERTLFGEGDSSLAVFETSVGRLLG 158  
DB 110 TSLVDRSGRIVGKYRKVHLPGHKEPQGRKHQHLEKRYFEPGD-LGFGVWRAFDGVMGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEEIHCAAW--PSFLYPNAKALGPDVNVAAASRIYAVEGQC 216  
DB 169 CICNDRWPETRYVMGLQGVEMVMLGYNTPYDHTGHDIDSLTQFHNHLSMQAGAYQNST 228  
QY 217 FVLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYA-- 274  
DB 229 WVIQTAK-----CGTEEGSKMV---GQSVIVAP-SGEIVAMACTIEDEIITARC 273  
QY 275 NLDPGVRIAKMAADPAGHYSRPDITRLLIDRSPKLP 311  
DB 274 DLDMGKR-YRETIFFDFAH-REPDAYRLIVERKGAVP 308

RESULT 15  
US-08-294-871A-46  
Sequence 46, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:

APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Nanba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiko





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:01:20 ; Search time 109.044 Seconds  
(without alignments)  
1291.301 Million cell updates/sec

Title: US-09-751-299-4

Perfect score: 1765

Sequence: 1 MKEAIKVACVQAAPLYMDLE.....DLRPYALGKASSETGAQLLEEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	100.0	337	3 US-09-751-299-4	Sequence 4, Appli
2	1761	99.8	337	4 US-10-146-772-386	Sequence 386, App
3	1761	99.8	337	4 US-10-241-742-386	Sequence 386, App
4	1761	99.8	337	4 US-10-440-523-386	Sequence 386, App
5	1761	99.8	337	4 US-10-440-503-386	Sequence 386, App
6	1761	99.8	337	4 US-10-461-925-386	Sequence 386, App
7	1339	75.9	337	4 US-10-146-772-48	Sequence 58, Appl
8	1339	75.9	337	4 US-10-146-772-58	Sequence 58, Appl
9	1339	75.9	337	4 US-10-241-742-48	Sequence 48, Appl
10	1339	75.9	337	4 US-10-241-742-58	Sequence 58, Appl
11	1339	75.9	337	4 US-10-440-523-48	Sequence 48, Appl
12	1339	75.9	337	4 US-10-440-523-58	Sequence 58, Appl
13	1339	75.9	337	4 US-10-440-503-48	Sequence 48, Appl
14	1339	75.9	337	4 US-10-440-503-58	Sequence 58, Appl
15	1339	75.9	337	4 US-10-461-925-48	Sequence 48, Appl
16	1339	75.9	337	4 US-10-461-925-58	Sequence 58, Appl
17	1139	64.5	333	4 US-10-146-772-292	Sequence 292, App
18	1139	64.5	333	4 US-10-146-772-292	Sequence 292, App
19	1139	64.5	333	4 US-10-440-523-292	Sequence 292, App
20	1139	64.5	333	4 US-10-440-503-292	Sequence 292, App
21	1139	64.5	333	4 US-10-461-925-292	Sequence 292, App
22	1136	64.4	332	4 US-10-146-772-140	Sequence 140, App
23	1136	64.4	332	4 US-10-241-742-140	Sequence 140, App
24	1136	64.4	332	4 US-10-440-523-140	Sequence 140, App
25	1136	64.4	332	4 US-10-440-503-140	Sequence 140, App
26	1136	64.4	332	4 US-10-461-925-140	Sequence 140, App
27	1091	61.8	355	4 US-10-146-772-306	Sequence 306, App

28	1091	61.8	355	4 US-10-241-742-306	Sequence 306, App
29	1091	61.8	355	4 US-10-440-523-306	Sequence 306, App
30	1091	61.8	355	4 US-10-440-503-306	Sequence 306, App
31	1091	61.8	355	4 US-10-461-925-306	Sequence 306, App
32	1081	61.2	345	4 US-10-146-772-334	Sequence 334, App
33	1081	61.2	345	4 US-10-241-742-334	Sequence 334, App
34	1081	61.2	345	4 US-10-440-523-334	Sequence 334, App
35	1081	61.2	345	4 US-10-440-503-334	Sequence 334, App
36	1081	61.2	345	4 US-10-461-925-334	Sequence 334, App
37	1051	59.5	334	4 US-10-146-772-6	Sequence 6, Appli
38	1051	59.5	334	4 US-10-241-742-6	Sequence 6, Appli
39	1051	59.5	334	4 US-10-440-523-6	Sequence 6, Appli
40	1051	59.5	334	4 US-10-440-503-6	Sequence 6, Appli
41	1051	59.5	334	4 US-10-461-925-6	Sequence 6, Appli
42	1042	59.0	353	4 US-10-146-772-212	Sequence 212, App
43	1042	59.0	353	4 US-10-241-742-212	Sequence 212, App
44	1042	59.0	353	4 US-10-440-523-212	Sequence 212, App
45	1042	59.0	353	4 US-10-440-503-212	Sequence 212, App

ALIGNMENTS

RESULT 1  
US-09-751-299-4  
; Sequence 4, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; APPLICANT: Chaplin, Jennifer A.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; PRIOR FILING DATE: 2000-12-29  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
; OTHER INFORMATION: environmental sample  
US-09-751-299-4

Query Match 100.0%; Score 1765; DB 3; Length 337;  
Best Local Similarity 100.0%; Pred. No. 1.3e-176;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKEAIKVACVQAAPLYMDLEATVDKTEIEMBEAARNMARLIAFPETWIPGYPWLMDSP	60
DB	1	MKEAIKVACVQAAPLYMDLEATVDKTEIEMBEAARNMARLIAFPETWIPGYPWLMDSP	60
QY	61	AWAMQFVROYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGTLVISQWFIGDNGD	120
DB	61	AWAMQFVROYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGTLVISQWFIGDNGD	120
QY	121	TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI	180
DB	121	TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI	180
QY	181	HCAAMPFSFLYPNAKALGPDVNVAASTRIVAVEGQCFVLASCALVSQSMIDMLCTDDEKH	240
DB	181	HCAAMPFSFLYPNAKALGPDVNVAASTRIVAVEGQCFVLASCALVSQSMIDMLCTDDEKH	240
QY	241	ALLLAGGHSRIIGPDGDLVAPLAENBEGILYANLDPGVRIAKMAADPAGHSRPDIT	300
DB	241	ALLLAGGHSRIIGPDGDLVAPLAENBEGILYANLDPGVRIAKMAADPAGHSRPDIT	300

Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300

QY 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337  
Db 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 2

US-10-146-772-386  
; Sequence 386, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: NitriIaees  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for windows version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-386

Query Match 99.8%; Score 1761; DB 4; Length 337;  
Best Local Similarity 99.7%; Pred. No. 3.5e-176;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVACVQAPRIYMDLEATVDKTIEMEEAARNARLIAFPETWIPGYPMFLWLDSP 60  
Db 1 MKEAIKVACVQAPRIYMDLKATVDKTIEMEEAARNARLIAFPETWIPGYPMFLWLDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTILYISQWFIGDNGD 120  
Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTILYISQWFIGDNGD 120  
QY 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
QY 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337  
Db 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 3

US-10-241-742-386  
; Sequence 386, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: NitriIaees  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for windows version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-386

Query Match 99.8%; Score 1761; DB 4; Length 337;  
Best Local Similarity 99.7%; Pred. No. 3.5e-176;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVACVQAPRIYMDLEATVDKTIEMEEAARNARLIAFPETWIPGYPMFLWLDSP 60  
Db 1 MKEAIKVACVQAPRIYMDLKATVDKTIEMEEAARNARLIAFPETWIPGYPMFLWLDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTILYISQWFIGDNGD 120  
Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTILYISQWFIGDNGD 120  
QY 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
QY 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337  
Db 301 RLLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 4

```
US-10-440-523-386
; Sequence 386, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitriases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-386

Query Match          99.8%; Score 1761; DB 4; length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAPITYMDLEATVDKTIEMEEAARNNARLIAFPETWIPGYWFLWLDSP 60
    |||||
Db 1 MKEAIKVCVQAPITYMDLKATVDKTIEMEEAARNNARLIAFPETWIPGYWFLWLDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
    |||||
Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
    |||||
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
    |||||
Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADBPAGHSRPDIT 300
    |||||
Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADBPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337
    |||||
Db 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 5
US-10-440-503-386
; Sequence 386, Application US/10440503
; Publication No. US20040038419A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-386

Query Match          99.8%; Score 1761; DB 4; length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAPITYMDLEATVDKTIEMEEAARNNARLIAFPETWIPGYWFLWLDSP 60
    |||||
Db 1 MKEAIKVCVQAPITYMDLKATVDKTIEMEEAARNNARLIAFPETWIPGYWFLWLDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
    |||||
Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
    |||||
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
    |||||
Db 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADBPAGHSRPDIT 300
    |||||
Db 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADBPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337
    |||||
Db 301 RLIDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 6
US-10-461-925-386
; Sequence 386, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
```



```
/ SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-386

Query Match      99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKEAIKVCVQAAPITMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLMLDSP 60
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1 MKEAIKVCVQAAPITMDLKATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLMLDSP 60

QY      61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120

QY      121 TIGARRKLKPTFVERTLFEGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      121 TIGARRKLKPTFVERTLFEGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180

QY      181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240

QY      241 ALLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      241 ALLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300

QY      301 RLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      301 RLIDRSPKLPVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 7
US-10-146-772-48
; Sequence 48, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
```

```
/ FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-48

Query Match      75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9.1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY      1 MKEAIKVCVQAAPITMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWFLMLDSP 60
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      1 VKEAIKVCVQAAPVFLDLDAITVDKTIALIEAARNGARLIAFPETWIPGYPWFLMLDSP 60

QY      61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      61 AWAMQFVRQYHENSLELDGPSQAKRISDAQAGIYVALGYSERVSGTLYMGQWLIDDKGE 120

QY      121 TIGARRKLKPTFVERTLFEGEGDSSLAVFETSVGRLLGGLCCWEHLQPLTKYALYAQNEEI 180
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      121 TAGLRRLKPTFVERTLFEGEGDSSLSTFDITPLGVLLGGLCCWEHLQPLSKYALYAQNEEI 180

QY      181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      181 HFAAMPFSIYRQATEVLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMICTDESKH 240

QY      241 ALLAGGHSRIIGPDGDLVAPLAENEEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
        :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      241 SLQAGGYSRIIGPDGDLARPLGENEGLYATLDPAARIYAKTAADPAGHSRPDVT 300

QY      301 RLIDRSPKLPVEIEGDLRPYALGKASETGAQLE 335
        |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB      301 RLINRSANQPVVEVGREI-----PASAQFEVE 329

RESULT 8
US-10-146-772-58
; Sequence 58, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-58

Query Match      75.9%; Score 1339; DB 4; Length 337;
```

Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY      1 MKEAIKVACVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWLWDSP 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 VKEAIKVACVQAAPVFLDLDTVDKTVALLIEEARNGARLIAFPETWIPGYPWLWDSP 60
QY      61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGTLTYSOWFIGDNGD 120
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 AWGMQFVRRYHENSFLVLDSPQAKRISDAQAGIYVALGYSERVSGLTYMGQWLIDDKGE 120
QY      121 TIGARRKLTPTFVERTLFGEGDSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 TAGLRRLKPTHTVERTLFGEGDSSSLSTFDTPLGVLGGLCCWEHLQPLSKYALYAQNEEI 180
QY      181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 HFAAMPFSFIYRQATEVLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMLTDESKH 240
QY      241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      241 SLQAGGGYSRIIGPDGSDLARPLGENEBGILYATLDPAAARIYAKTAADPAGHSRPDVT 300
QY      301 RLLIDRSPKLPVEIEGDLRPYALGKASSETGAQLE 335
      ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      301 RLLINRSANQPVVEVGREI-----PASAQGFVE 329
```

## RESULT 9

```
US-10-241-742-48
; Sequence 48, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
;   APPLICANT: Short, Jay
;   APPLICANT: Weiner, David
;   APPLICANT: Chaplin, Jennifer
;   APPLICANT: Chi, Ellen
;   APPLICANT: Milan, Aileen
;   APPLICANT: Desantis, Grace
;   APPLICANT: Madden, Mark
;   TITLE OF INVENTION: Nitriases
;   FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
;   CURRENT APPLICATION NUMBER: US/10/241,742
;   PRIOR FILING DATE: 2002-09-09
;   PRIOR APPLICATION NUMBER: US/10/146,772
;   PRIOR FILING DATE: 2002-05-15
;   PRIOR APPLICATION NUMBER: US 60/309,006
;   PRIOR FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US 60/351,336
;   PRIOR FILING DATE: 2002-01-22
;   PRIOR APPLICATION NUMBER: US 60/300,189
;   PRIOR FILING DATE: 2001-06-21
;   PRIOR APPLICATION NUMBER: US 09/751,299
;   PRIOR FILING DATE: 2000-12-28
;   PRIOR APPLICATION NUMBER: US 60/254,414
;   PRIOR FILING DATE: 2000-12-07
;   PRIOR APPLICATION NUMBER: US 60/173,609
;   PRIOR FILING DATE: 1999-12-29
;   NUMBER OF SEQ ID NOS: 386
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 48
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-48
```

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY      1 MKEAIKVACVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWLWDSP 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 VKEAIKVACVQAAPVFLDLDTVDKTVALLIEEARNGARLIAFPETWIPGYPWLWDSP 60
QY      61 AWAMQFVRQYHENSLELDGPQAKRISDAAKRLGIMVTLGMSERVGTLTYSOWFIGDNGD 120
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 AWGMQFVRRYHENSFLVLDSPQAKRISDAQAGIYVALGYSERVSGLTYMGQWLIDDKGE 120
QY      121 TIGARRKLTPTFVERTLFGEGDSSSLAVFETSVGRLGGLCCWEHLQPLTKYALYAQNEEI 180
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 TAGLRRLKPTHTVERTLFGEGDSSSLSTFDTPLGVLGGLCCWEHLQPLSKYALYAQNEEI 180
QY      181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
      ||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 HFAAMPFSFIYRQATEVLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMLTDESKH 240
QY      241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      241 SLQAGGGYSRIIGPDGSDLARPLGENEBGILYATLDPAAARIYAKTAADPAGHSRPDVT 300
QY      301 RLLIDRSPKLPVEIEGDLRPYALGKASSETGAQLE 335
      ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      301 RLLINRSANQPVVEVGREI-----PASAQGFVE 329
```

## RESULT 10

```
US-10-241-742-58
; Sequence 58, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
;   APPLICANT: Short, Jay
;   APPLICANT: Weiner, David
;   APPLICANT: Chaplin, Jennifer
;   APPLICANT: Chi, Ellen
;   APPLICANT: Milan, Aileen
;   APPLICANT: Desantis, Grace
;   APPLICANT: Madden, Mark
;   APPLICANT: Burk, Mark
;   TITLE OF INVENTION: Nitriases
;   FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
;   CURRENT APPLICATION NUMBER: US/10/241,742
;   PRIOR FILING DATE: 2002-09-09
;   PRIOR APPLICATION NUMBER: US/10/146,772
;   PRIOR FILING DATE: 2002-05-15
;   PRIOR APPLICATION NUMBER: US 60/309,006
;   PRIOR FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US 60/351,336
;   PRIOR FILING DATE: 2002-01-22
;   PRIOR APPLICATION NUMBER: US 60/300,189
;   PRIOR FILING DATE: 2001-06-21
;   PRIOR APPLICATION NUMBER: US 09/751,299
;   PRIOR FILING DATE: 2000-12-28
;   PRIOR APPLICATION NUMBER: US 60/254,414
;   PRIOR FILING DATE: 2000-12-07
;   PRIOR APPLICATION NUMBER: US 60/173,609
;   PRIOR FILING DATE: 1999-12-29
;   NUMBER OF SEQ ID NOS: 386
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 58
;   LENGTH: 337
;   TYPE: PRT
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-58
```

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY      1 MKEAIKVACVQAAPITYMDLEATVDKTIELMEEAARNNARLIAFPETWIPGYPWLWDSP 60
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 VKEAIKVACVQAAPVFLDLDTVDKTVALLIEEARNGARLIAFPETWIPGYPWLWDSP 60
```

```

QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGLTYSQWFIGDNGD 120
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AWGMQFVRRYHENSLELDSPQAKRISDAQAGIYVALGYSERVSGTLYMGQWLIDDKGE 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEI 180
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 TAGLRRLKPTHTVERTLFGEGDSSLSLTFDPLGVGLCCWEHLQPLSKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HFAAMPFSFIYRQATEVLGPEVNVAASRIYAVEGQCFVLASCALVSPEMIEMLCCTDESKH 240
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRDPIT 300
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 SLQAGGGYSRIIGPDGSDLARPLGENEGILYATLDPARIRYAKTAADPAGHSRDPVT 300
QY 301 RLIDRSPKL PVEIEGDLRPPYALGKASETGAOLE 335
    ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 RLINRSANQPVVEVGREI-----PASAQGFVE 329

```

```

RESULT 11
US-10-440-523-48
; Sequence 48, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-48

```

```

Query Match 75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9.1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVCVQAAPITMDLEATVDKTIELMEEAARNARLIAFPETWIPGYWFLWLDSP 60
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VKEAIKVCVQAAPVFLDLDATVDKTVALLIEAARNGARLIAFPETWIPGYWFLWLDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGLTYSQWFIGDNGD 120
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 61 AWGMQFVRRYHENSLELDSPQAKRISDAQAGIYVALGYSERVSGTLYMGQWLIDDKGE 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEI 180
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 TAGLRRLKPTHTVERTLFGEGDSSLSLTFDPLGVGLCCWEHLQPLSKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HFAAMPFSFIYRQATEVLGPEVNVAASRIYAVEGQCFVLASCALVSPEMIEMLCCTDESKH 240
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRDPIT 300
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 SLQAGGGYSRIIGPDGSDLARPLGENEGILYATLDPARIRYAKTAADPAGHSRDPVT 300
QY 301 RLIDRSPKL PVEIEGDLRPPYALGKASETGAOLE 335
    ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 RLINRSANQPVVEVGREI-----PASAQGFVE 329

```

```

RESULT 12
US-10-440-523-58
; Sequence 58, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for windows Version 4.0
; SEQ ID NO 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-58

```

```

Query Match 75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9.1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVCVQAAPITMDLEATVDKTIELMEEAARNARLIAFPETWIPGYWFLWLDSP 60
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VKEAIKVCVQAAPVFLDLDATVDKTVALLIEAARNGARLIAFPETWIPGYWFLWLDSP 60
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGLTYSQWFIGDNGD 120
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AWGMQFVRRYHENSLELDSPQAKRISDAQAGIYVALGYSERVSGTLYMGQWLIDDKGE 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEI 180
    |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
Db      121 TAGLRRLKPTHTVERTLFGEGDSSLSLTFDTPLGLGGLCCWEHLQPLSKYALYAONEI 180
QY      181 HCAAMPSSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
Db      181 HFAAMPSSFSIRQATEVTLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMLCTDESKH 240
QY      241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300
Db      241 SLQAGGGYSRIIGPDGSDLARPLGENEGILYATLDPAAARIYAKTAADPAGHSRPDVT 300
QY      301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLE 335
Db      301 RLINRSANQPVEVEGREI-----PASAQGFVE 329
```

RESULT 13  
US-10-440-503-48  
; Sequence 48, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440,503  
; PRIOR FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for windows Version 4.0  
; SEQ ID NO 48  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-48

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;  
QY 1 MKEAIKVCQAAPRYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPFWLMDSP 60  
Db 1 VKEAIKVCQAAPRYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPFWLMDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQDNGD 120  
Db 61 AWGMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQDNGD 120  
QY 121 TIGARRKLPKPTHTVERTLFGEGDSSLSLTFDTPLGLGGLCCWEHLQPLSKYALYAONEI 180  
Db 121 TAGLRRLKPTHTVERTLFGEGDSSLSLTFDTPLGLGGLCCWEHLQPLSKYALYAONEI 180  
QY 181 HCAAMPSSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HFAAMPSSFSIRQATEVTLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMLCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300  
Db 241 SLQAGGGYSRIIGPDGSDLARPLGENEGILYATLDPAAARIYAKTAADPAGHSRPDVT 300  
QY 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLE 335  
Db 301 RLINRSANQPVEVEGREI-----PASAQGFVE 329

RESULT 14  
US-10-440-503-58  
; Sequence 58, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark J.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440,503  
; PRIOR FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380,737  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for windows Version 4.0  
; SEQ ID NO 58  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-58

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;  
QY 1 MKEAIKVCQAAPRYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPFWLMDSP 60  
Db 1 VKEAIKVCQAAPRYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPFWLMDSP 60  
QY 61 AWAMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQDNGD 120  
Db 61 AWGMQFVRQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQDNGD 120  
QY 121 TIGARRKLPKPTHTVERTLFGEGDSSLSLTFDTPLGLGGLCCWEHLQPLSKYALYAONEI 180  
Db 121 TAGLRRLKPTHTVERTLFGEGDSSLSLTFDTPLGLGGLCCWEHLQPLSKYALYAONEI 180  
QY 181 HCAAMPSSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HFAAMPSSFSIRQATEVTLGPEVNVAAASRIYAVEGQCFVLASCALVSPEMIEMLCTDESKH 240  
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAGHSRPDIT 300  
Db 241 SLQAGGGYSRIIGPDGSDLARPLGENEGILYATLDPAAARIYAKTAADPAGHSRPDVT 300  
QY 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLE 335  
Db 301 RLINRSANQPVEVEGREI-----PASAQGFVE 329

RESULT 15  
US-10-461-925-48  
; Sequence 48, Application US/10461925  
; Publication No. US20040053378A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001  
; CURRENT APPLICATION NUMBER: US/10/461,925



```
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-48
```

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY 1 MKEAIKVCVQAAPVIMDLVDTKTIEMEEAARNNARLIAFPETWIPGYPMFLWLDSP 60
Db 1 VKEAIKVCVQAAPVIMDLVDTKTIEMEEAARNNARLIAFPETWIPGYPMFLWLDSP 60
QY 61 AWAMQFVROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
Db 61 AWGMOFVRRYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120
QY 121 TIGARRKCLKPTFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEBI 180
Db 121 TAGLRKCLKPTFVERTLFGEGDSSSLSTFDTPPLGLGLCCWEHLQPLSKYALYAQNEBI 180
QY 181 HCAAMPSSFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
Db 181 HFAAMPSSFSLYRQATEVLGPEVNVAAASRIYAVEGQCFVLASCALVSPENIEMLCCTDESKH 240
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENEIGILYANLDPGVRIILAKMAADDPAGHYSRPDIT 300
Db 241 SLQAGGGSRIIGPDGDLARPLGENEEGILYATLDPAAARIYAKTAADDPAGHYSRPDVT 300
QY 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAQLE 335
Db 301 RLINRSANQPEVEVGREI-----PASAQGFVE 329
```

Search completed: April 27, 2006, 01:06:08  
Job time : 109.044 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:02:36 ; Search time 17.7628 Seconds  
(without alignments)  
862.816 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVCVQAAPITMDLE.....DLRPVALGKASGTGAQLLEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New: \*  
1: /SIDS5/ptodata/2/pubpaa/US08\_NEW\_PUB.pep: \*  
2: /SIDS5/ptodata/2/pubpaa/US06\_NEW\_PUB.pep: \*  
3: /SIDS5/ptodata/2/pubpaa/US07\_NEW\_PUB.pep: \*  
4: /SIDS5/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep: \*  
5: /SIDS5/ptodata/2/pubpaa/US09\_NEW\_PUB.pep: \*  
6: /SIDS5/ptodata/2/pubpaa/US10\_NEW\_PUB.pep: \*  
7: /SIDS5/ptodata/2/pubpaa/US11\_NEW\_PUB.pep: \*  
8: /SIDS5/ptodata/2/pubpaa/US60\_NEW\_PUB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	822	46.6	356	US-10-537-075-7	Sequence 7, Appli
2	780	44.2	369	US-10-919-182-18	Sequence 18, Appl
3	777	44.0	369	US-10-919-182-16	Sequence 16, Appl
4	776	44.0	369	US-10-919-182-4	Sequence 4, Appli
5	776	44.0	369	US-10-919-182-6	Sequence 6, Appli
6	774	43.9	369	US-10-919-182-12	Sequence 12, Appl
7	774	43.9	369	US-10-919-182-14	Sequence 14, Appl
8	765	43.3	369	US-10-919-182-8	Sequence 8, Appli
9	451	25.6	333	US-11-096-568A-20687	Sequence 20687, A
10	451	25.6	351	US-11-096-568A-20686	Sequence 20686, A
11	446.5	25.3	330	US-11-096-568A-6955	Sequence 6955, Ap
12	446.5	25.3	350	US-11-096-568A-6954	Sequence 6954, Ap
13	323	18.3	233	US-11-096-568A-6956	Sequence 6956, Ap
14	274.5	15.6	193	US-11-096-568A-20688	Sequence 20688, A
15	125	7.1	295	US-11-079-463-5457	Sequence 5457, Ap
16	114.5	6.5	259	US-11-045-004-822	Sequence 822, App
17	109	6.2	296	US-11-045-004-2100	Sequence 2100, Ap
18	106	6.0	460	US-11-087-099-8682	Sequence 8682, Ap
19	97	5.5	388	US-10-497-135-19	Sequence 19, Appl
20	97	5.5	388	US-11-269-215-19	Sequence 19, Appl
21	95	5.4	252	US-11-055-822-134	Sequence 134, App
22	95	5.4	252	US-11-055-822-1120	Sequence 1120, Ap
23	92	5.2	153	US-10-986-501-173	Sequence 173, App
24	92	5.2	386	US-10-497-135-20	Sequence 20, Appl
25	92	5.2	386	US-11-269-215-20	Sequence 20, Appl

26	88	5.0	463	7	US-11-087-099-5333	Sequence 5333, Ap
27	88	5.0	489	7	US-11-188-298-3840	Sequence 3840, Ap
28	86	4.9	358	7	US-11-087-099-5821	Sequence 5821, Ap
29	86	4.9	358	7	US-11-188-298-16344	Sequence 16344, A
30	85	4.8	487	7	US-11-188-298-12261	Sequence 12261, A
31	84	4.8	304	6	US-10-467-657-7410	Sequence 7410, Ap
32	84	4.8	369	7	US-11-188-298-15842	Sequence 15842, A
33	83	4.7	451	7	US-11-188-298-21140	Sequence 21140, A
34	82.5	4.7	3655	7	US-11-075-185-5	Sequence 5, Appli
35	82	4.6	356	7	US-11-188-298-8290	Sequence 8290, Ap
36	81.5	4.6	509	7	US-11-188-298-4316	Sequence 4316, Ap
37	81.5	4.6	1084	7	US-11-079-463-5327	Sequence 5327, Ap
38	81	4.6	455	7	US-11-188-298-11004	Sequence 11004, A
39	81	4.6	500	7	US-11-188-298-6835	Sequence 6835, Ap
40	81	4.6	585	7	US-11-045-004-1585	Sequence 1585, Ap
41	81	4.6	1065	6	US-10-511-989-28	Sequence 28, Appl
42	80.5	4.6	290	7	US-11-188-298-1550	Sequence 1550, Ap
43	80.5	4.6	306	7	US-11-188-298-17485	Sequence 17485, A
44	80.5	4.6	474	7	US-11-188-298-2797	Sequence 2797, Ap
45	80.5	4.6	475	7	US-11-087-099-2233	Sequence 2233, Ap

ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zelinski, Thomas  
; APPLICANT: Hauer, Bernhard  
; TITLE OF INVENTION: L-RHAMNOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537,075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
; US-10-537-075-7

Query Match 46.6%; Score 822; DB 6; Length 356;  
Best Local Similarity 48.2%; Pred. No. 2.8e-71;  
Matches 160; Conservative 56; Mismatches 114; Indels 2; Gaps 1;

QY	2	KEAIKVCVQAAPITMDLEATVDKTIELMEBARNNARLIAFPETWIPGYPMFLDSPA	61
DB	4	KIVRAAAVQAASPNYDLATGVDKTIELARQAARDEGCDLIVGETWLPGYPFHWLGAPA	63
QY	62	WAMQFVRQYHENSLELDGPOAKRISDAKRLGIWVTLGMSERVGTLTYSQWFIGDNGDT	121
DB	64	WSLKYSARYYANGSLDSAEFQRIQAARTLGIFALGYSESRSGSLYLGCCLIDDKGEM	123
QY	122	IGARRKLKPTFVERTLFGEGDGSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEIHH	181
DB	124	LMSRRKLKPTHTVERTYFGEGYARDLIVSDTELGRVGALCCWEHLSPLSKYALYQGEHALH	183
QY	182	CAAMPSPSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHA	241
DB	184	IAAMPSPSLYSEQAHALSAKVNMMAASQIYSVEGOCFTIAASSVTQETLDMLEVGHNAP	243
QY	242	LILAGGHSRIIGPDGDLVAPLAENEBGILYANLDPGVRILAKMAADPAGHSRPDITR	301
DB	244	LLKVGGSMSMIFAPDGRTIAPYLPHPDAEGLIADLNMEIEIAFAKAIINDPVGHYSKPEATR	303



[illegible]

```

RESULT 5
US-10-919-182-6
; Sequence 6, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B2 and H9
US-10-919-182-6

```

Query Match	44.0%;	Score 776;	DB 6;	Length 369;
Best Local Similarity	47.6%;	Pred. No. 7.8e-67;		
Matches 150;	Conservative 61;	Mismatches 100;	Indels 4;	Gaps 3
QY	8	ACVQAPDIYMDLEATVDKTIELMEEAARNARLIAFPETWIPGYPFWLWIDSPAWAQFV	67	
DB	11	ATVQAEFPWLADADATIDKSIGIIEEAQKASLIAPEVFIPGYPYMAWLGDVKYSLSFT	70	
QY	68	ROYHENSLELDGPOAKRISDAAKRLGIMVTLGMSERVGTLYISQWFIGDNGDTIGARRK	127	
DB	71	SRYHENSLELGDDRMRLQLAARRNKIALVMGYSEREAGSRYLSQVFIDERGEIVANRRK	130	
QY	128	LKPTFVERTLFGEGDSSIAVFETSVGRLLGCLCWEHLQPLTYKALYAQNEETHCAWPS	187	
DB	131	LKPTHVERTIYGEENGDTFLTHDFAFGRVGLNCWEHFQPLSKFPMYSLGEQYHVASWPA	190	
QY	188	FS-LYPNAKALGPDVNVAASTRIVAVEGOCFVLASCALVSQSMIDMLCTDDEKHALLAG	246	
DB	191	MSPLQPDVFQ-LSTEANATVARSYALEGOTFVLCTQVIGPSAIEFTCLNDEQFALLPQG	249	
QY	247	GGHSRIIGPDGDLVAPLAENEGLLYANLDPGVRILAKMAADPAGHSRDPITRLLIDR	306	
DB	250	CGWARIYGPDSSELAKPLAEDAEGLIYAIEDLEQILLAKAGADPVGHYSRDPVLSVQFDP	309	
QY	307	SPKLPV--VEIEGDL	319	
DB	310	RNHTPVHRIGIDGRL	324	

RESULT 6  
US-10-919-182-12  
; Sequence 12, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. dupont de Nemours and Company, Inc  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark

```

; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
US-10-919-182-12

```

[illegible]

```

RESULT 7
US-10-919-182-14
; Sequence 14, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
US-10-919-182-14

```

Query Match	43.9%	Score 774;	DB 6;	Length 369;
Best Local Similarity	47.6%	Pred. No. 1.2e-66;		
Matches	150;	Conservative 61;	Mismatches 100;	Indels 4;
				Gaps 3;

QY 8 ACVOAPITYMDLEATVDKTIELMEEARARNARLIAFPETWIPGYPFWLMLDSPAMAMQFV 67



Db 11 ATVOAEPVWLADATIDKSIGIIEEAQKASLIAPEVFIPGYPYWAWLGDVKYSLSFT 70

QY 68 ROYHENSLELDGPQAKRISDAKRLGIMVTGMSERVGGTLTYSQWFIGDNGDTIGARRK 127

Db 71 SRYHENSLELDGDDRMRLQLAARRNKIALVMGYSEREAGSRYLQVFI DERGEIVANRRK 130

QY 128 LKPTFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEIHCAWPS 187

Db 131 LKPTHVERTIYGEGNGTDFLTHDFAFGRVGGLNCWEHKQPLSKEMMYSLGEQVHVASWPA 190

QY 188 FS-LYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALLAG 246

Db 191 MSPLQPDVFQ-LSIEANATVTRSYAIEGQFVLCSTQVIGPSAIEFTFCUNDEQALLPQG 249

QY 247 GGHSRITGPDGDLVAPLAENEGLYANLDPGVRILAKMAADPAGHSRDPDITRLIDR 306

Db 250 CGWARIYGPDGSELAKEPLAEDAEGLVYAEIDLEQIILLAKAGADPVGHYSRDPVLSVQFDP 309

QY 307 SPKLPV--VEIEGDL 319

Db 310 RNHTPVHRIIGIDGRL 324

RESULT 8

US-10-919-182-8

; Sequence 8, Application US/10919182

; Publication No. US20060035352A1

; GENERAL INFORMATION:

; APPLICANT: E.I. duPont de Nemours and Company, Inc.

; APPLICANT: Di Cosimo, Robert

; APPLICANT: Payne, Mark

; APPLICANT: O'Keefe, Daniel

; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS

; FILE REFERENCE: CL2584 US NA

; CURRENT APPLICATION NUMBER: US/10/919,182

; CURRENT FILING DATE: 2004-08-16

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 8

; LENGTH: 369

; TYPE: PRT

; ORGANISM: Artificial sequence

; FEATURE:

; OTHER INFORMATION: Mutant nitrilase B4

; US-10-919-182-8

Query Match 43.3%; Score 765; DB 6; length 369;

Best Local Similarity 47.6%; Pred. No. 9e-66;

Matches 150; Conservative 58; Mismatches 103; Indels 4; Gaps 3;

QY 8 ACVOAARIYMDLEATVDKTIELMEEAARNNARLIAPEWTIPGYPFWLWLDSPAWAMQFV 67

Db 11 ATVOAEPVWLADATIDKSIGIIEEAQKASLIAPEVFIPGYPYWAWLGDVKCSLSFT 70

QY 68 ROYHENSLELDGPQAKRISDAKRLGIMVTGMSERVGGTLTYSQWFIGDNGDTIGARRK 127

Db 71 SRYHENSLELDGDDRMRLQLAARRNKIALVMGYSEREAGSRYLQVFI DERGEIVANRRK 130

QY 128 LKPTFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEEIHCAWPS 187

Db 131 LKPTHVERTIYGEGNGTDFLTHDFAFGRVGGLNCWEHKQPLSKIMMYSLGEQVHVASWPA 190

QY 188 FS-LYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALLAG 246

Db 191 MSPLQPDVFQ-LSIEANATVTRSYAIEGQFVLCSTQVIGPSAIEFTFCUNDEQALLPQG 249

QY 247 GGHSRITGPDGDLVAPLAENEGLYANLDPGVRILAKMAADPAGHSRDPDITRLIDR 306

Db 250 CGWARIYGPDGSELAKEPLAEDAEGLVYAEIDLEQIILLAKAGADPVGHYSRDPVLSVQFDP 309

QY 307 SPKLPV--VEIEGDL 319

Db 310 RNHTPVHRIIGIDGRL 324

RESULT 9

US-11-096-568A-20687

; Sequence 20687, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nickolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; TITLE OF INVENTION: Theryby

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 20687

; LENGTH: 333

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(333)

; OTHER INFORMATION: Ceres Seq. ID no. 12386686

US-11-096-568A-20687

Query Match 25.6%; Score 451; DB 7; length 333;

Best Local Similarity 33.3%; Pred. No. 1.4e-35;

Matches 107; Conservative 68; Mismatches 118; Indels 28; Gaps 8;

QY 6 KVAQOAPRIYMDLEATVDKTIELMEEAARNNARLIAPEWTIPGYP-----WFLWIDSP 60

Db 13 RVTVOQASSVFYDTPATLIDKAEKLVAAAGYGSQLVLPFEVFGYPHGSTFGLVGNRT 72

QY 61 AWAMQFVROYHENSLELDGPQAKRISDAKRLGIMVTGMSERVGGTLTYSQWFIGDNGD 120

Db 73 AKGKEDFQKYHASAIDVPGEVGRSLALAGKYKVLVIGVERAGYTLXNTVLSFDPGK 132

QY 121 TIGARRKLPFVERTLFGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAQNEE 180

Db 133 YLGKHKRKMPTALERVFWGFGDGTIPYDTPIGKMGALICWENRMPPLRTAMYAKGIEI 192

QY 181 HCAWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMID-----MLCT 235

Db 193 YCA--PYVDCMPTW-----LSSMTHIALLEGCFVLSACQFCRKKNYPPPEYTFPG 241

QY 236 DDEK---HALLLAGGHSRITGPDGDLVAPLAENEGLYANLDPGVRILAKMAADPAG 292

Db 242 LEEEPSPEVVCSSG--SVTISPLGTVLAGPNYES-EALLTADLDLGEIVRAKFDVVG 298

QY 293 HYSRDPDITRLIDRSPKLPV 313

Db 299 HYSRPEVLSLVKSDPK-PAV 318

RESULT 10

US-11-096-568A-20686

; Sequence 20686, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nickolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 20686

; LENGTH: 351

; TYPE: PRT

; ORGANISM: Zea mays subsp. mays

; FEATURE:

; NAME/KEY: misc\_feature

; LOCATION: (1)..(351)

; OTHER INFORMATION: Ceres Seq. ID no. 12386685

US-11-096-568A-20686

Query Match 25.6%; Score 451; DB 7; Length 351;  
Best Local Similarity 33.3%; Pred. No. 1.5e-35;  
Matches 107; Conservative 68; Mismatches 118; Indels 28; Gaps 8;

QY 6 KVACVQAAPITYMDLEATVDKTIELMEBAARNARLIAFPETWIPGYP-----WFLMLDSP 60  
DB 31 RVTVVQASSVFYDTPATLDKAEXLVAAAGVGSQVLFPFVFGVGYPHGSTFGLVGNRT 90  
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIGDNGD 120  
DB 91 AKGKEDFQKYTHASAIIDVPGPEVSRLSALAGKYKVFVIGVERAGTYLNTVLSFDPLGK 150  
QY 121 TIGARRKLKPTFVERTLFEGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALYAONEI 180  
DB 151 YLGKHKRKWPTALERVWFEGDSTIPYDPIGKMGAICWENRMPPLRTAMYAKGIEI 210  
QY 181 HCAWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFLVASCALVQSMID----MLCT 235  
DB 211 YCA--PTVDCWPTW-----LSSMTHIALEGCFVLSACQFCRRKNYPPPEYTFCG 259  
QY 236 DDEK--HALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAG 292  
DB 260 LEEBPSPESTVCSG--SVIISPLGTVLAGPNYES-EALLTADLDLGEIVRAKFDVVG 316  
QY 293 HYSRPDITRLIDRSPKLPV 313  
DB 317 HYSRPEVLSLVKSDPK-PAV 336

## RESULT 11

US-11-096-568A-6955  
; Sequence 6955, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Thierby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6955  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(330)  
; OTHER INFORMATION: Ceres Seq. ID no. 15169318  
US-11-096-568A-6955

Query Match 25.3%; Score 446.5; DB 7; Length 330;  
Best Local Similarity 33.5%; Pred. No. 3.8e-35;  
Matches 111; Conservative 60; Mismatches 113; Indels 47; Gaps 10;

QY 5 IKVACVQAAPITYMDLEATVDKTIELMEBAARNARLIAFPETWIPGYPFWFLMLDSPAWAM 64  
DB 10 VRATVQASTIFYDTPATLDKAERLIAEATSYGSQLVFPFAFVGYP-----RGSAGFL 64  
QY 65 QF-----VRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWF 114  
DB 65 SIGNRTVKGREERKXTHASAIIDVPGPEVDRILAAAGKYKXHLVMGVIERDGYTLCTVLF 124  
QY 115 IGDNGDTIGARRKLKPTFVERTLFEGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALY 174  
DB 125 FDSQGHYLGKHKRKIMPTALERVIWFGDSTIPVFETPVGKIGALICWENRMPPLRTAMY 184  
QY 175 AQNEETHCAWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFLVLA---CA----- 223  
DB 185 AKGVEIYCA-----PTAD--ARDVQASMTHTIALEGCFVLSANQFCRRRDYPPPP 233

QY 224 --LVQSMDMLCTDEKHALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVR 281  
DB 234 EYVFAGTEVDL--TPD--SVVCAGG--SVIISPLGAVLAGPNYDG-EALISADLDLGEI 285  
QY 282 ILAKMAADPAGHYSRPDITRLIDRSPKLPV 312  
DB 286 ARAKFDPDVVGHYSRPEVLSLTVKDHPTNPV 316

## RESULT 12

US-11-096-568A-6954  
; Sequence 6954, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Thierby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6954  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(350)  
; OTHER INFORMATION: Ceres Seq. ID no. 15169317  
US-11-096-568A-6954

Query Match 25.3%; Score 446.5; DB 7; Length 350;  
Best Local Similarity 33.5%; Pred. No. 4.1e-35;  
Matches 111; Conservative 60; Mismatches 113; Indels 47; Gaps 10;

QY 5 IKVACVQAAPITYMDLEATVDKTIELMEBAARNARLIAFPETWIPGYPFWFLMLDSPAWAM 64  
DB 30 VRATVQASTIFYDTPATLDKAERLIAEATSYGSQLVFPFAFVGYP-----RGSAGFL 84  
QY 65 QF-----VRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWF 114  
DB 85 SIGNRTVKGREERKXTHASAIIDVPGPEVDRILAAAGKYKXHLVMGVIERDGYTLCTVLF 144  
QY 115 IGDNGDTIGARRKLKPTFVERTLFEGEGDSSSLAVFETSVGRGLCCWEHLQPLTKYALY 174  
DB 145 FDSQGHYLGKHKRKIMPTALERVIWFGDSTIPVFETPVGKIGALICWENRMPPLRTAMY 204  
QY 175 AQNEETHCAWPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFLVLA---CA----- 223  
DB 205 AKGVEIYCA-----PTAD--ARDVQASMTHTIALEGCFVLSANQFCRRRDYPPPP 253  
QY 224 --LVQSMDMLCTDEKHALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVR 281  
DB 254 EYVFAGTEVDL--TPD--SVVCAGG--SVIISPLGAVLAGPNYDG-EALISADLDLGEI 305  
QY 282 ILAKMAADPAGHYSRPDITRLIDRSPKLPV 312  
DB 306 ARAKFDPDVVGHYSRPEVLSLTVKDHPTNPV 336

## RESULT 13

US-11-096-568A-6956  
; Sequence 6956, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Thierby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6956

```

;      LENGTH: 233
;      TYPE: PRT
;      ORGANISM: Glycine max
;      FEATURE:
;      NAME/KEY: misc_feature
;      LOCATION: (1)..(233)
;      OTHER INFORMATION: Ceres Seq. ID no. 15169319
;
US-11-096-568A-6956

```

Query Match	18.3%	Score 323;	DB 7;	Length 233;
Best Local Similarity	35.9%	Pred. No. 1.9e-23;		
Matches	85;	Conservative	39;	Mismatches 81;
				Indels 32;
				Gaps 8;

```

QY      89 AKRLGIMWTIGMSERVGGTLYISQWFIQNDGTIGARRKLKPTVERTLFGEGDSSLAY 148
      | : : : | | | | : | | | : | | | : | | | : |
Db      2 AGKYKVLVWGVIERDGYTLCTVLFQDSQGHYLGKHKRKIMPTALERVIWFGDGSITPV 61

```

```
QY      149 PETSVGRLGGLCCWEHLQP LTKYALYAONEEIHCAAMPSSFLYNAAKALGPDVNVAASR   208
        |||::|||:|||:|||:|||::|
Db       62 PETPVGKIGAIICWENRMPLLR TAMYAKGEIYCA-----PTAD---ARDVWGQSMT    110
```

```

QY      209 IYAVEGQCFTLAS---CA-----LVQSMDIMLCTDDEKALLLAGGHSRIIG 255
      |||||:::|:::|:::|:::|:::|
Db      111 HIALEGGCFVLASNQFCRRRDYPPPEYVFACTEVDL--TPD---SYVCAG--SVIISF 163

```

QY 256 DGGDLVAPLAENEGILYANLDFGVRI LAKMAADPAGHSRPDI TRLLIDRSKPLEY 312

| | : ||| | | : | |

Db 164 LGAVLAGPNYDG -EALISADLDGEIARAKFD FVVVGHSRPEVL SLTVKDHPTNPV 219

```

RESULT 14
US-11-096-568A-20688
; Sequence 20688, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Therby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20688
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(193)
; OTHER INFORMATION: Ceres Seq. ID no. 12386687
US-11-096-568A-20688

```

Query Match	15.6%;	Score 274.5;	DB 7;	Length 193;
Best Local Similarity	35.4%;	Pred. No. 7e-19;		
Matches 68;	Conservative 36;	Mismatches 65;	Indels 23;	Gaps 7;

```
QY      130 PTFVERTLFGEGDGSSIAVFETSVGRLLGGLCCWEHTLOPLTKALYAONEEIHCAAMPSPS 189
```

```
Db      2 PTALERVFWGFGDGSTIPYDTPIGKMGALLICWENRMPILRTAMYAKGIEIYYCA--PTVD 59
```

```

QY      190 LYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMID-----MLCTDEK---HA 241
      |      ::  |::| |::|  :  ::  :
Db      60 CMTPTW-----LSSMTHIALBGGGCFVLASACQFCRRKNYPPEPYTFCGLBEEPSPE 110

```

```
QY      242 LLLAGGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIIAKMAADPAHGYSRPPDITR 301
       :: :||| | ||| | | | : | : | : || | | | : :
Db     111 VVCSGG--SVTIISPLGTIVLAGNYTES-EALLTADLDLGEIVAKEDFDVGHSRPEVL 167
```

```

QY      302 LLIDRSPKLPV 313
          |::|||
Db      168 LVVKSDPK-PAV 178

```

```

RESULT 15
US-11-079-463-5457
; Sequence 5457, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRAG
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5457
; LENGTH: 295
; TYPE: PRT
; ORGANISM: B.fragilis
US-11-079-463-5457

```

Query Match	7.1%;	Score 125;	DB 7;	Length 295;
Best Local Similarity	22.7%;	Pred. No. 0.00032;		
Matches	78;	Conservative 44;	Mismatches 106;	Indels 116;
				Gaps 19;

```

QY      5 IKVACVQAPIYMDLEATVDKTIEM-----EEAARNNAALLIAPETWIPGYPMFLMD 58
      ||| : ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5 IKVGIIQQA-----NTSDIRINMLNLAKSIEACAANGAHLVVLQE-----LH 46

```

```

QY      59 SPAMAMQFVRQYHENSLEL---DGEQAKRISD--AAKRLGIMVTLLGMSERVGGLYISQ 112
      :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db     47 N---SLYFCQCTENTDLFELAEPIPGSPSTGYSELAAANRI-VLVTSLFEKRAPGLYHNTA 102

```

```

Qy      113 WFIGDNGDTIGARRKL----KPTFVERTLFGEGDGSSLAVFETSVGRLLGGLCCWEHTLOPL 168
      : | | | | : | | | : | | | | | :
Db      103 VFEDRDGSLAGKRYRKMHIPDDPAYVEKRYFTPGD-IGFEPIQTSIGKLGVLVCWDQ---- 157

```

QY 169 TKYALYAQNBEIHCAAMPSESLYPNNAKALGPDVNVAASTRIVAEQCFVLASCALVSQS 228  
 158 -----W-----YPEAA-----RLMALKGAEILITYPTAIGWES 184  
 Db

```

QY      229 MIDMLCTDEK---HALLLAGGHSRIIC-----PDGGDLVAPLAENEIGIL 274
          ||||| : :: : | : :
Db      185 ---TDTDEKKQLNAMIISQRAHAAVANGLEPVSIVNRVGHPEPD-----PSGQTNGILF 23

```

```

QY      273  YANLDPGVRLAKMAADPAGHSRPDITRLRIDRSKPLVEIE 316
      : | | | | : : | | : : | | :
Db      235  WGN-----SFVAGPQGEY----LAQAGNDRSENM-IYEVD 264

```

Search completed: April 27, 2006, 01:06:51  
Job time : 18.7628 secs

```

RESULT 15
US-11-079-463-5457
; Sequence 5457, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRAG
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5457
; LENGTH: 295
; TYPE: PRT
; ORGANISM: B.fragilis
US-11-079-463-5457

```

Query Match	7.1%;	Score 125;	DB 7;	Length 295;
Best Local Similarity	22.7%;	Pred. No. 0.00032;		
Matches	78;	Conservative 44;	Mismatches 106;	Indels 116;
				Gaps 19;

```

QY      5 IKVACVQAPIYMDLEATVDKTIEM-----EEAARNNAALLIAPETWIPGYPMFLMD 58
      ||| : ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      5 IKVGIIQQA-----NTSDIRINMLNLAKSIEACAANGAHLVVLQE-----LH 46

```

```

QY      59 SPAMAMQFVRQYHENSLEL---DGEQAKRISD--AAKRLGIMVTLLGMSERVGGLYISQ 112
      :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db     47 N---SLYFCQCTENTDLFELAEPIPGSPSTGYSELAAANRI-VLVTSLFEKRAPGLYHNTA 102

```

```

Qy      113 WFIGDNGDTIGARRKL----KPTFVERTLFGEGDGSSLAVFETSVGRLLGGLCCWEHTLOPL 168
      : | | | | : | | | : | | | | | :
Db      103 VFEDRDGSLAGKRYRKMHIPDDPAYVEKRYFTPGD-IGFEPIQTSIGKLGVLVCWDQ---- 157

```

QY 169 TKYALYAQNBEIHCAAMPSESLYPNNAKALGPDVNVAASTRIVAEQCFVLASCALVSQS 228  
 158 -----W-----YPEAA-----RLMALKGAEILITYPTAIGWES 184  
 Db

```

QY      229 MIDMLCTDEK---HALLLAGGHSRIIC-----PDGGDLVAPLAENEIGIL 274
          ||||| : :: : | : :
Db      185 ---TDTDEKKQLNAMIISQRAHAAVANGLEPVSIVNRVGHPEPD-----PSGQTNGILF 23

```

```

QY      273  YANLDPGVRLAKMAADPAGHSRPDITRLRIDRSKPLVEIE 316
      : | | | | : : | | : : | | :
Db      235  WGN-----SFVAGPQGEY----LAQAGNDRSENM-IYEVD 264

```

Search completed: April 27, 2006, 01:06:51  
Job time : 18.7628 secs